JUMS

Volume 9, Issue 3, September 2024

JUNIOR MANAGEMENT SCIENCE

Published by Junior Management Science e.V.

ISSN: 2942-1861

Junior Management Science 9(3) (2024) 1591-1608

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Lost in Transcription: Experimental Findings on Ethnic and Age Biases in AI Systems

Anna Smerdiagina

Technical University of Munich

Abstract

Artificial intelligence (AI) has revolutionized industries and improved our lives in various ways. However, AI systems' potential to amplify existing biases in society has become a major concern. This thesis explores the concept of bias in AI and how it can lead to discrimination, focusing specifically on the performance of Automatic Speech Recognition (ASR) systems in relation to the ethnicity (accent) of participants. The study collected 187 recordings from proficient English speakers of 55 ethnic groups.These recordings were transcribed via ASR systems and evaluated by the word error rate (WER) metric. The ASR systems selected for the study were Gboard (Android) by Google, Apple keyboard (iOS), and Whisper by Open AI. The study results show that ethnicity significantly impacts the performance of ASR systems, with some ethnic groups experiencing substantially higher error rates than others. The study provides evidence that ASR systems may not be equally accurate for all users. To address ethnic bias in AI systems, it is essential to take a multi-faceted approach involving technical and societal solutions. The findings highlight the importance of addressing bias in AI systems to ensure fairness, transparency, and equity for all users, regardless of ethnicity.

Keywords: automatic speech recognition; bias in AI; digital ageism; digital equity; ethnic bias

1. Introduction

Artificial intelligence (AI) is revolutionizing many industries and making our lives easier in various ways (Bostrom, 2014). However, bias in AI systems has become a major concern in recent years due to the potential of perpetuating and amplifying biases that are already present in society. Bias in AI can occur at various stages of the development process, including the selection of training data, the design of the algorithms, and the interpretation of the results (Mehrabi et al., 2021). One particular area of concern is the potential for AI systems to exhibit biases based on ethnicity and age, which can result in discrimination against certain groups of people (Barocas & Selbst, 2016).

The impact of bias in AI can be significant, as these systems are increasingly being used in a variety of contexts, including hiring, lending, and criminal justice (Kleinberg et al., 2018). For example, a biased AI system that is used in the hiring process may unfairly reject job candidates from certain ethnicities or age groups. Similarly, a biased AI system that is used in the criminal justice system may disproportionately affect certain groups of people, leading to further inequalities and injustices.

In this thesis, we will explore the concept of bias in AI and how it can lead to discrimination. We will review the literature on the ways in which AI systems can exhibit biases based on ethnicity and age. We will consider the impact that these biases can have on individuals and society as a whole, and explore potential ways to mitigate them. Further, we will conduct an experiment to investigate the extent to which these biases exist in specific AI systems.

The hypothesis of this thesis is that the performance of automatic speech recognition (ASR) systems may be influenced by the ethnicity (accent) of participants. We will test this

I sincerely thank Prof. Dr. Isabell Welpe, my academic guide, whose unwavering trust in my creative journey and agile supervision has been an invaluable compass. My heartfelt thanks to the 210 participants whose voices enriched the exploration of fairness in automatic speech recognition. Special appreciation goes to Edman Paes dos Anjos for his constant support and belief in me.

DOI: https://doi.org/10.5282/jums/v9i3pp1591-1608

hypothesis on three ASR systems: Gboard by Google, Apple keyboard by Apple Inc., and Whisper by OpenAI, and compare the results of native English speakers with the results of non-native English speakers from various ethnic groups.

The research questions of this thesis are as follows:

1. How does the bias in AI systems affect automatic speech recognition for different ethnic groups?
2. Does the ethnicity (accent) of a user affect the performance of automatic speech recognition systems?
3. What are the best practices to reduce bias in AI systems, specifically in the context of ethnicity and age?

Overall, this thesis aims at contributing to the understanding of biases in AI and providing insights on how they can be addressed in order to create more equitable AI systems. We hope to raise awareness of the potential consequences of bias in AI for vulnerable groups and provide recommendations for addressing biases in the development and deployment of AI systems.

The thesis is structured as follows: we will start by providing a theoretical background on bias in AI, including different types of biases and how they can occur. We will then examine the literature on ethnicity and age bias in AI systems. Further, we will discuss measures for reducing these biases in AI systems. The methodology chapter will describe the experiment we conducted to investigate the extent to which these biases exist in the specific AI system. Finally, the results and discussion chapter will present and analyze the results of the experiment and discuss their implications.

2. Definition of Bias

At any moment in time, there is a stream of 11 billion bits of information coming to us from every sense that we have. The human mind is only able to consciously process 40 bits (Zimmermann, 1986). This implies that most of our decisions are subconscious or unconscious. Given the overwhelming amount of information, the nervous system is only able to function through the use of cognitive shortcuts, also known as heuristics (Storage, 2021). However, these shortcuts can instigate discriminative behavior.

In this chapter, we will examine the types of biases that can manifest in AI systems, including cognitive biases, biases in machine learning, and biases in AI speech recognition. We will also explore the sequence of stereotypes, prejudice, and discrimination and how it leads to biased outcomes. The main focus of this chapter is to provide a comprehensive understanding of the various forms of biases that can manifest in AI systems and differentiate key definitions.

2.1. Cognitive Bias and Its Implications

It is essential to differentiate the key definitions to understand where and how biases arise. The concept of "cognitive misers," or the tendency to rely on mental shortcuts when making decisions due to limited cognitive resources, can contribute to biases (Fiske & Taylor, 1991). These mental shortcuts can lead to the automatic activation of stereotypes and biases, causing biased decisions and outcomes (Kleinberg, 2018).

Cognitive biases can manifest through the sequence of a stereotype, prejudice, and discrimination (Dasgupta & Asgari, 2004; Storage, 2021). Stereotypes are beliefs about the characteristics of a group of people, and these beliefs can be based on a variety of factors such as race, gender, sexual orientation, and religion (Dovidio, 2001). Stereotypes can be either positive or negative, and can lead to negative attitudes, or prejudice, towards certain groups (Dovidio, 2001; Storage, 2021). For instance, the stereotype that men are more capable and competent in the workplace than women can lead to prejudice towards women in the workplace, resulting in discriminatory behaviors such as not promoting women to leadership positions or paying them less than their male counterparts (Eagly & Karau, 2002).

In contrast with a stereotype, prejudice can only be negative (Storage, 2021). Prejudice, or negative attitudes towards a group of people, can then manifest in discriminatory behavior towards a certain group (Kleinberg, 2018). For instance, prejudice against individuals with disabilities might lead to discrimination, such as denying employment or education opportunities (Olkin & Pledger, 2003). The sequence of stereotypes, prejudice, and discrimination can aggravate existing societal biases and disparities (Barocas & Selbst, 2016).

One of the ways that biases can manifest is through explicit and implicit biases (Greenwald & Krieger, 2006). Explicit biases are conscious and intentional, and people may be aware of their own explicit biases (Greenwald et al., 2015). For instance, a study showed that Uber and Lyft drivers were canceling rides or extending wait times for African–American customers based on their names and faces upon the order, which is a direct and intentional form of discrimination (Ge et al., 2016). Additionally, the study found that women were taken on longer routes to extend the cost of the fare, also a direct indication of explicit bias. According to Lee (2018), explicit biases can and must be traced and mitigated further by law reinforcement.

Implicit biases, on the other hand, are unconscious and automatic (Greenwald et al., 2015). Implicit biases can be particularly insidious as they are not always recognized or acknowledged, yet they can still influence behavior (Nosek et al., 2002). This case is illustrated in an experiment that showed that using word embeddings in machine learning processes can lead to sexist results (Bolukbasi et al., 2016). For instance, in word analogy tests, "man" would be assigned to "computer programmer" while "woman" would be assigned to "homemaker." This bias toward women triggered the authors to propose a method that respects the embeddings for gender-specific words but de-biases embeddings for gender-neutral words.

In conclusion, cognitive biases in AI systems have the potential to exacerbate existing societal issues, causing biased outcomes. Understanding the different types of biases that can occur in machine learning is crucial in developing effec-

tive strategies for mitigating these biases. The next section will delve deeper into these specific types of biases in AI systems.

2.2. Types of Biases in Machine Learning

Machine learning, as a subfield of artificial intelligence, has become an integral part of a human routine, from food delivery to airport security procedures, affecting every individual in various ways (Guegan & Hassani, 2018; Guimaraes & Tofighi, 2018). However, one of the major challenges facing machine learning is the presence of biases in the data that is used to train these models, as well as flawed training and testing processes. These biases can lead to unfair and inaccurate outcomes, particularly for marginalized groups. In this section, we will explore the different types of biases that can occur in machine learning and the methods that can be used to address them.

2.2.1. Data bias

Data bias refers to the systematic errors or distortions that occur when the data used to train or evaluate machine learning models is unrepresentative or skewed in some way (Baeza-Yates, 2018). Data bias can be caused at any phase in a range of areas, from human reporting and selection bias to annotator bias (Hellström et al., 2020). The use of AI systems that are trained on biased data has the potential to amplify harmful stereotypes about certain ethnicities. For instance, an AI system trained on data that includes negative stereotypes about certain ethnicities may influence the way individuals are treated or perceived, escalating inequality.

2.2.2. Sampling bias

In the field of machine learning, sampling bias occurs when the sample of data used to train a machine learning model is not representative of the population it is intended to model (Mehrabi et al., 2021). If a model is trained on data that is predominantly from one gender or race, it may not accurately reflect the characteristics of the broader population and may lead to biased results. In 2018, Reuters reported that an AI system used to evaluate job applicants by Amazon's Human Resources department was biased to advise hiring male candidates, resulting in fewer female individuals being offered job opportunities (Dastin, 2018).

2.2.3. Selection bias

Selection bias occurs when the data used to train a model is selected in a non-random manner, resulting in a sample that is not representative of the population (Shah et al., 2020). This can occur when data is self-selected, such as in online surveys, or when data is selected based on certain criteria, such as data from only a particular geographic region (Baeza-Yates & Ribeiro-Neto, 2011).

2.2.4. Measurement bias

Measurement bias refers to errors or distortions in the way data is collected, recorded, or measured (Suresh & Guttag, 2019). For instance, if data is collected using a biased survey instrument or by a researcher with a preconceived notion about the outcome, the resulting data may be biased (Hajian et al., 2016).

2.2.5. Label bias

Label bias, or annotator bias (Hellström et al., 2020), refers to inconsistent labeling processes: when different annotators have mismatching styles that lead to misunderstanding and get reflected in the labels created. A common occurrence of label biases happens when differing labels get assigned to the same type of object by different annotators (for instance, grass vs. lawn, painting vs. picture) (Malisiewicz & Efros, 2008).

2.2.6. Confirmation bias

Confirmation bias is a type of cognitive bias that occurs when people seek out or interpret information in a way that confirms their preexisting hypotheses or opinions. In the context of machine learning, confirmation bias can occur when data is selected or analyzed in a way that confirms the researcher's expectations or hypotheses, leading to partial results (Carvalho et al., 2019). Some researchers recognize confirmation bias as a sub-type of a label bias (Srinivasan & Chander, 2021).

2.2.7. Negative Set bias

Negative set bias refers to the unreasonable emphasis on negative examples (examples that the model is attempting to classify as a particular class) in comparison to positive examples (examples that are not being classified as that particular class). As a result, datasets that only collect data on negative instances might be biased and disadvantaged due to poor modeling of the rest of the visual world (Torralba & Efros, 2011).

For example, in the context of email classification, if the training dataset includes a higher proportion of spam emails than non-spam emails, the machine learning model may be more sensitive to spam emails and may classify a higher proportion of non-spam emails as spam (Zhou et al., 2014).

Negative set bias can be mitigated by balancing the training dataset or weighting the training data to give greater importance to positive examples (Chawla, 2005).

2.2.8. Problem Framing bias

Problem framing errors can also cause bias (Srinivasan & Chander, 2021). For instance, if a credit card company aims at predicting customer trustability using AI, the concept of creditworthiness must be well-defined and estimated. However, "creditworthiness" is a rather vague concept (Barocas & Selbst, 2016). Problem framing strongly depends on the company's goals: maximizing the profit margin or maximizing the number of repaid loans.

However, as Solon Barocas, an assistant professor at Cornell University who specializes in fairness in machine learning emphasizes, "those decisions are made for various business reasons other than fairness or discrimination" (Hao, 2019). If the algorithm discovered that granting subprime loans lead to profit maximization, it would eventually lead to predatory behavior, even if it was not the intention of the company.

2.2.9. Recent Bias Mitigation Approaches

It is essential to be aware of these types of data bias and build a versatile mitigation strategy in order to avoid their effects and ensure that machine learning models are accurate and reliable (Baeza-Yates & Ribeiro-Neto, 2011). In order to address data bias in machine learning, it is recommended to use diverse and representative datasets, apply statistical techniques to adjust for bias, and use multiple methods to validate results (Suresh & Guttag, 2019).

Additionally, the use of human-in-the-loop approaches, where a human is involved in the decision-making process, can also help to mitigate bias in AI systems (Xin et al., 2018). However, some studies warn that systems with one or too few human experts are insufficient due to human agent's bias. One solution to it might be a hybrid pipeline with multiple human experts and a classifier to share the decision making load and reduce bias (Keswani et al., 2022).

It is also important to be transparent about the data sources and methods used in order to allow for external scrutiny and reproducibility (Baeza-Yates & Ribeiro-Neto, 2011).

3. Digital Ageism in AI systems

The digital age has brought numerous advancements and innovations that have transformed the way we live, work, and communicate. However, these advancements have also led to the emergence of a new form of discrimination known as digital ageism (Hunsaker & Hargittai, 2018). Nowadays, digital ageism is addressed as a critical issue and a global priority by the World Health Organization (WHO) in their annual Global Report on Ageism (World Health Organization, 2022).

Digital ageism refers to the discrimination or prejudice against individuals based on their age or generation in the digital world. Digital ageism can manifest in various ways, such as the exclusion of older individuals from technology training and education, the assumption that older individuals are not capable of using technology, and the creation of age-based stereotypes in the media and advertising (Charles & Carstensen, 2010; Zickuhr & Smith, 2012).

In this chapter, we will examine the various forms of digital ageism and how they impact older individuals in the digital world. We will also discuss the ways in which digital ageism intersects with other forms of bias, such as racial and ethnic bias, and how these intersections can compound and amplify the negative effects on marginalized groups. Finally, we will explore potential solutions for addressing and combating digital ageism in order to create a more inclusive and equitable digital society for all.

3.1. Forms of Digital Ageism

Digital ageism encompasses a range of forms of discrimination, including exclusion from technology training and education, negative stereotypes and prejudices, and lack of accessibility of technology for older adults.

One common form of digital ageism is the exclusion of older individuals from technology training and education (Czaja et al., 2008). This can occur when older individuals are not offered the same opportunities for technology training and education as their younger counterparts, leading to a lack of digital literacy and skills among older adults. The consequences of exclusion from technology can be significant for older adults: from limited access to job opportunities and social connections, to contribution to social isolation (Hultsch et al., 1999).

Another form of digital ageism is the assumption that older individuals are not capable of using technology (Choi et al., 2020; Palmore, 2001). Such a stereotype can lead to older individuals being excluded from certain technological platforms and experiences, or being treated with condescension when attempting to use technology. Older individuals who rely on technology can be particularly disadvantaged in their daily activities, such as staying in touch with loved ones or managing their health.

Older adults often face discrimination during the design process of digital technologies. Such evidence is presented in a recent study that analyzed 7 facial image datasets. Age discrimination was manifested in the labeling of the datasets, where extensive age intervals were assigned to older adults in datasets (Chu et al., 2022). For instance, groups for participants of younger age were categorized into narrow age groups within each dataset, such as 13 to 19, and 20 to 36 years old, compared to a considerably more pervasive category 60+ or 66+ years old, despite decades of physical and mental changes for those individuals.

Digital ageism can also manifest in the form of age-based stereotypes and prejudices in the media and advertising (de Paula Couto & Wentura, 2017). For example, older individuals may be depicted as out-of-touch or unable to keep up with new technologies, leading to negative stereotypes that can further exclude them from participating in the digital world (de Paula Couto & Wentura, 2017).

Overall, there is doubtfully enough data to represent older individuals. Essentially, the existing data also fails to include and depict healthy ageing, underrepresenting older adults' needs, interests, and aspirations, which confirms ageist stereotypes (Chu et al., 2022).

3.2. Intersections of Digital Ageism and Other Forms of Bias

Digital ageism often intersects with other forms of bias, compounding and amplifying the negative effects of bias on marginalized groups (Drydakis et al., 2018; World Health Organization, 2022).

For instance, older individuals from marginalized racial and ethnic groups may face double discrimination due to both their age and their racial or ethnic identity (Drydakis et al., 2018). The study shows that older applicants received a lower number of job interview invitations compared to younger participants. However, the study also states that a study group with people of color as participants had even worse vacancy access. The outcome implies that people with minority ethnicities face a higher level of ageism compared to the majority race representatives (Drydakis et al., 2018).

In addition, digital ageism can intersect with other forms of bias in the development and design of technology. AI and machine learning algorithms that are trained on biased datasets may produce biased outputs that disproportionately negatively impact certain age and racial or ethnic groups (Mehrabi et al., 2021). This can occur in a plethora of contexts, such as in the development of age-based or racially biased advertising or the use of AI in hiring decisions (Mehrabi et al., 2021).

A recent study shows that the intersection of ageism and sexism is a prominent combination even among designers and developers of technologies for older people (Chen & Petrie, 2022). The authors conducted a qualitative study with in-depth interviews with technology designers and developers and found that both male and female participants held negative attitudes toward older workers. In particular, older women were found to face double discrimination due to the intersection of ageism and sexism. This is consistent with previous studies that have shown that women face intersectional discrimination based on their race and gender, in addition to age (Harnois, 2014; Stypińska, 2021).

These intersections of digital ageism and other forms of bias can have significant negative impacts on marginalized groups, such as limiting access to job opportunities and social connections and contributing to social isolation and decreased social capital. It is important to recognize and address these intersections in order to create a more inclusive and equitable digital society for all individuals, regardless of age or identity (Drydakis et al., 2018).

3.3. Potential Solutions

There are several potential solutions for addressing and combating digital ageism. One widely suggested approach is to increase the availability and accessibility of technology training and education for older individuals (Friemel, 2016; Mitzner et al., 2010; Niehaves & Plattfaut, 2014). This solution involves providing targeted technology training programs for older adults, as well as ensuring that these programs are available in a variety of locations and formats to accommodate different learning styles and needs (Mitzner et al., 2010).

Apart from accessibility, a positive user experience (UX) can play a significant role in bringing safety and comfort to older adults as users of various applications. A study shows that the UX in information and communication technologies, poorly adjusted to older individuals' needs and user behavior patterns, distances them from the digital world, causing digital exclusion and, consequently, feeling of loneliness among the participants (Lagacé et al., 2015). A recent study by Chen and Petrie (2022) reaffirms that technology specialists for older people as users should receive adequate de-biasing training in order to reduce the number of biased experts in the field.

Another solution is to challenge and debunk age-based stereotypes and prejudices in the media and advertising (Zickuhr & Smith, 2012). The specific steps can be promoting more positive and accurate portrayals of older individuals in the media, as well as calling out and addressing instances of ageism in advertising and media content.

Overall, there is a need for more inclusive and equitable design and development of technology, including AI and machine learning algorithms. It is crucial to ensure that these systems are trained on diverse and representative datasets, as well as implementing measures to mitigate and address potential biases in the outputs of these systems (Mehrabi et al., 2021).

3.4. Successful Practices and Initiatives Worldwide

It is worth mentioning the initiatives and programs that have been implemented to combat ageist digital inequalities worldwide. For instance, the European Commission's (EAEA) campaign "New skills agenda for Europe" (2019) aims to promote the development of digital skills, including those of older adults, and ensure that they are not left behind in the digital transformation. The "Silver Surfers" program in the UK, launched by Age UK and TalkTalk, provides training and support for older adults to help them acquire the digital skills they need to participate fully in the digital world (Age UK & TalkTalk, 2014).

Similarly, US-based non-profit organizations, such as American Association of Retired Persons (AARP), are focusing on issues affecting older individuals over age fifty. As of 2018, the AARP group reported to have made significant contributions towards improving the lives of over 38 million members, including providing access to better economic security, consumer protection, and healthcare, promoting affordability and quality in long-term care, and fostering the development of livable communities (AARP, n.d.). Their program (Older Adult Technology Services (OATS), 2022), provides resources and support for older adults to learn about and engage with technology, breaking down barriers to digital inclusion.

Initiatives aimed at promoting digital inclusion for older adults are highly relevant to the ageing population in Japan, which is one of the fastest-growing in the world. According to the Annual Report on the Aging Society (2017), dementia is forecasted to have an effect on one in five people in Japan by 2025. Access to digital skills and resources is vital for older individuals in Japan to participate fully in modern society and maintain their quality of life.

There are several notable initiatives in Japan aimed at promoting digital inclusion for older adults. For instance, the

Ministry of Internal Affairs and Communications launched the "Silver Human Resources Center" program in 1974, to support older job seekers (Weiss et al., 2005). Nowadays, Silver Human Resources Center also provides digital literacy training to older adults in Japan. The program aims to create a network of people who can support older adults in learning about and using digital technology.

In addition, the Japanese government has implemented policies to encourage businesses to develop age-friendly technologies and services, such as the "Universal Design" policy, which promotes the design of products and services that are accessible to all, regardless of age or ability (Ministry of Economy, Trade and Industry, 2020).

These initiatives demonstrate the Japanese government's commitment to promoting digital inclusion and addressing digital ageism in Japan, and serve as an important model for other countries to follow. As a result of these programs, unique cases have emerged, such as that of an 83-year-old female app game developer (Government of Japan, 2018) or a restaurant that is run and maintained by people affected by dementia (Government of Japan, 2019).

Raising awareness and a better understanding of digital ageism and its impacts on older individuals can facilitate a shift toward a more inclusive and equitable digital society for all, as well as increase life quality for older individuals.

4. Ethnic Bias in AI Systems

Ethnic bias in AI refers to the tendency for AI systems to produce biased outcomes that disproportionately harm or discriminate against certain racial or ethnic groups. Bias can occur when the data used to train AI systems reflects and reinforces existing societal biases and inequalities. Resolving ethnic bias in AI is vital due to far-reaching consequences for those who are targeted by it, including discrimination, marginalization, and reduced opportunities (Zafar et al., 2017). Additionally, ethnic bias in AI can magnify existing societal inequalities, leading to further harm and injustice (Bolukbasi et al., 2016).

In this chapter, we will examine the ways in which ethnicity-based bias can occur in AI, the consequences of this bias, and efforts to mitigate or eliminate it. We will review relevant studies on the topic and discuss practices for designing and evaluating AI systems to reduce the risk of ethnicity-based bias.

4.1. Impact of Ethnic Bias in AI Systems

The issue of ethnic bias in artificial intelligence has gained increasing attention in recent years. Incidents have exposed the vulnerability of AI to perpetuating existing societal biases, leading to calls for increased efforts to identify and address such biases in order to promote justice and equality for all individuals.

One example of ethnic discrimination in the digital world is the use of targeted advertising by Facebook (now: Meta Platforms). In 2016, it was revealed that Facebook allowed advertisers to exclude African, Hispanic, and other "ethnic affinities" from seeing advertisements (Ali et al., 2019). Such practice magnified existing inequalities and discrimination, as individuals from certain ethnicities may have been disproportionately excluded from seeing certain advertisements based on their zip code or other factors that are correlated with ethnicity. Despite public exposure and repeated media investigations, the problem has remained over years (Ali et al., 2019; Angwin & Parris, 2016; Angwin et al., 2017).

Another study shows that in order to select a look-alike audience, Facebook tries to infer the attributes that distinguish the audience from the general population, recurrently causing representation bias. Such bias distribution might potentially incline biases in a source audience of several thousand to a lookalike audience of tens of millions (Speicher et al., 2018). As a result of this type of discrimination, individuals from certain ethnicities may be disadvantaged in terms of access to job opportunities, housing, and other resources (Dastin, 2018).

Finally, a lack of ethnic and racial diversity has been observed within academic settings (D. Zhang et al., 2021). According to the AI Index 2021 Annual Report, among the new AI PhDs in the USA in 2019, the largest percentage (45.6%) are white representatives (non-Hispanic), followed by Asian representatives (22.4%). In comparison, only 2.4% were African American (non-Hispanic) and 3.2% were of Hispanic ethnicities. Ethnic underrepresentation in AI research and development can limit the diversity of design and deployment of AI systems, highlighting the need for increased efforts to promote diversity and inclusivity in the field.

4.2. Consequences and Implications of Ethnic Bias

Ethnicity-based bias in AI can manifest in a variety of ways, including the amplification of existing societal biases and discrimination (Zafar et al., 2017). Some specific forms of ethnicity-based bias in AI include the issues depicted in this chapter.

4.2.1. Accuracy Disparities in Face Recognition

Accuracy disparities in AI systems can become a problem, as certain ethnicities may be more accurately represented in these systems, leading to unequal treatment depending on the ethnicities of individuals.

To illustrate, AI systems used for facial recognition or language processing may be more accurate for certain ethnicities, resulting in discriminative outcomes for individuals based on their ethnicities (Buolamwini & Gebru, 2018; Caliskan et al., 2017; Dastin, 2018). According to a recent study by MIT and Microsoft, false arrests or incorrect identification of suspects are possible as a result of poor recognition and identification accuracy by AI systems for individuals from certain ethnicities. The experiment by Buolamwini and Gebru (2018) shows that the maximum difference in face recognition error rate between the lighter skin tone male groups and darker skin tone female groups, best and worst classified groups respectively, is 34.4%. This accuracy disparity is largely caused by datasets exceedingly composed of

lighter-skinned participants (79.6% to 86.2%, depending on the dataset).

Additionally, another research has proved that these accuracy disparities can be particularly pronounced for individuals who are members of multiple marginalized groups, such as women of color (Else-Quest & Hyde, 2016).

4.2.2. Hiring and Lending

Implementation of AI in decision-making processes in hiring and lending has sparked concerns about the risk of ethnic bias and the exacerbation of current inequalities.

For instance, an AI system involved in the hiring process may be more likely to shortlist job applicants from certain ethnicities, causing an unfair advantage for individuals from those groups (Caliskan et al., 2017). A recent study showed that word embedding, a popular framework to transform text data into structured vectors that can be more easily processed by a computer, can lead to sexism and other forms of discrimination (Bolukbasi et al., 2016). Word embedding has been used in various machine learning tasks, including AI systems, trained to assist in hiring decisions.

Similarly, an AI system used in lending decisions may be more likely to approve loans for individuals from certain ethnic groups, leading to an unequal distribution of financial opportunities (Barocas & Selbst, 2016; Zafar et al., 2017).

4.2.3. Law Enforcement and Security Contexts

The use of AI systems in law enforcement and security contexts has garnered attention due to the potential for racial profiling and discrimination. AI systems used in these contexts may be more likely to identify individuals from certain ethnicities as potential suspects, leading to false accusations and other acts of discrimination (Caliskan et al., 2017; Else-Quest & Hyde, 2016; O'Neil, 2016).

In particular, there are several ways in which a predictive policing algorithm may impose discrimination on individuals of certain ethnicities. To illustrate, a predictive policing algorithm in Florida, United States, in 2013 and 2014 was more likely to falsely assign high risk scores to individuals of color as potential suspects, with only 20% of the correct prediction rate (Angwin et al., 2022). It is possible that the prediction algorithm was trained on biased data that includes a disproportionate number of individuals from certain ethnicities who have been arrested or convicted (Eubanks, 2018; O'Neil, 2016). If the algorithm is trained on this biased data, it may be more likely to identify individuals from those ethnicities as potential suspects, even if they are no more likely to commit crimes than individuals from other ethnicities.

Another possibility is that the algorithm is using factors correlated with ethnicity, such as zip code or socioeconomic status, as input (Eubanks, 2018). If these factors are correlated with ethnicity and are being used by the algorithm to predict the likelihood of criminal activity, it may be more likely to identify individuals from certain ethnicities as potential suspects, even if their ethnicity is not directly related to their likelihood of committing a crime (O'Neil, 2016).

Biased crime risk prediction systems can jeopardize the human rights of marginalized groups, leading to increased discrimination and social inequality. Therefore, it is crucial to implement strict regulations and conduct thorough checks before implementing these technologies.

4.3. Potential Solutions

Mitigation and elimination of these forms of bias should become vital in the development and deployment of AI systems. This can involve using diverse and representative datasets, implementing fairness and accountability measures, and regularly evaluating AI systems for bias (Dwork et al., 2012).

To address ethnic bias in AI systems, it is important to take a multi-faceted approach that involves both technical and societal solutions. One solution is the use of diverse and representative training data to develop AI algorithms in order to ensure that the algorithms are more representative of the populations they serve (Dwork et al., 2012).

Another solution is conducting regular bias audits of AI systems, and active monitoring of their performance for evidence of bias (Landers & Behrend, 2022). However, such audit and cross-checking imply additional hours of work for current employees, or, potentially, additional hiring. While large companies, such as IBM, can afford it and actively implement these complex techniques(Hobson & Dortch, 2022), startups and companies with exiguous budgets might not possess this opportunity. Alternatively, there are currently tools that help detect and measure bias in models, as well as calculate the bias drift over time (Simon, 2022). One such tool is Amazon SageMaker by Amazon Web Services (AWS).

In addition to these technical solutions, organizations must also prioritize diversity, equity, and inclusion in their hiring practices and organizational culture. This includes creating a diverse team of researchers and engineers and establishing ethical review processes for AI systems development (Floridi, 2019). A similar to suggested set of practices and tools is implemented within Re:work Unbiasing course, an open-source educational course by Google (2017). The course aims at reducing potential unconscious bias for hiring and promotion decisions by providing a set of practices, checklists, facilitator guides, and team discussion guides that can be adjusted to the user's team.

Implementing these solutions requires a commitment to transparency, accountability, and continuous improvement. This includes regularly reporting on the performance of AI systems and publishing data on their biases and limitations. Additionally, it is crucial for organizations to prioritize the development of ethics and governance frameworks to ensure the responsible development and deployment of AI systems (Floridi, 2019).

Addressing ethnic bias in AI systems is a complex and ongoing challenge. However, by taking a diversified approach, communities can help ensure that AI is developed and used in a way that is equitable, just, and beneficial for all.

5. Methodology

In this study, a controlled experiment was conducted to investigate the potential for biased results for different ethnic groups in ASR systems. The main objective of the study was to evaluate the performance of commonly used ASR systems in recognizing speech from non-native proficient English speakers of various ethnic groups, and put these results in comparison with canonical examples of American, Australian, and British native English speakers.

The methodology for this experiment included the selection of a sample of English-speaking participants representing a range of ethnicities, and the recording of speech samples from each participant. The speech samples were then manually checked in terms of quality standards fulfillment, and finally processed through the selected ASR systems. The ASR Systems selected for the study are as follows: Gboard for Android by Google, Apple keyboard for iOS, and Whisper by Open AI. The recognition accuracy was measured and compared across ethnic groups.

The study will provided insights into the potential biases in AI algorithms and their impact on different ethnic groups, which is discussed in the Discussion chapter of the thesis.

The experiment consists of four phases: Planning, Performing, Reviewing, and Closing stages, as shown in Figure 1.

The results of the study were analyzed using statistical methods to determine whether there are any significant differences in the recognition accuracy of the ASR systems for different ethnic groups.

This research aimed to contribute to the understanding of potential biases in AI algorithms and their impact on different ethnic groups and to identify possible solutions to mitigate such biases.

5.1. Automatic Speech Recognition Applications

According to recent data, the utilization of AI transcription applications has grown exponentially in recent years. Research has shown that the integration of AI in transcription applications has the potential to improve communication and accessibility for individuals with speech or language impairments (J. Zhang et al., 2023). The rapid growth in the use of these applications highlights the increasing importance of AI in this field.

Among the most widely used applications in this category are Gboard by Google, and the Apple keyboard, both are pre-installed on a large number of mobile devices nowadays. In contrast, another ASR selected for the experiment is a relatively new product picked to represent the state of the art of ASR solutions. Whisper by OpenAI has been described as "revolutionary" in the field of AI transcription by experts in the field (Ansari, 2022).

Gboard has emerged as a leading player in the field of AI transcription, with over 5 billion downloads reported in 2022 (Google Play, n.d.). The app is generally installed out-of-the-box on many Android-based mobile devices and supports over 900 languages, as per the application's description. Gboard is also available on the iOS operating system and can be installed on iOS-based devices through the App Store. However, both the number of downloads and supported languages are significantly lower on the iOS App Store (App Store, 2016).

The Apple iOS keyboard is a default application that is pre-installed on the iOS operating system. According to Tim Cook, Chief Executive Officer of Apple, there were at least 1.65 billion Apple devices, as of January 2021 (Nellis, 2021). This suggests a widespread availability and usage of the Apple iOS keyboard among individuals who utilize Apple devices.

The third voice transcription application under examination is Whisper, an open-source state-of-the-art ASR system. The system has been trained on a vast amount of supervised data, precisely, 680,000 hours of data according to Radford et al. (2022). Whisper is a research project developed by OpenAI, a leading AI research and deployment research company, based in California, USA. OpenAI was founded by notable figures in the technology industry, including Elon Musk and Sam Altman (OpenAI, 2022).

In contrast to Gboard and iOS virtual keyboards, is a large ASR model executed in clusters of servers in the cloud. As such, It is not meant for on-device real-time transcription like the other products surveyed. The ASR system takes an average of 25 to 140 seconds to transcribe a voice sample. This feature may be useful for more complex transcription tasks that require higher accuracy and can tolerate either long processing time or increased execution costs.

It is interesting to note that, while Gboard does not include punctuation in its transcription, Apple adds periods and commas based on pauses between words, and Whisper accurately places periods at the end of each sentence.

5.2. Participants and Data Collection

There are canonical samples of native speakers' recordings for Harvard sentences located in the Open Speech Repository and further used for comparison with non-native English-speaking participants' transcription data (Open Speech Repository, n.d.).

The study population included 210 participants, of which 23 participants' recordings were excluded from the experiment due to low audio quality or reading mistakes. The 187 recordings were collected from proficient and fluent English speakers of 55 different ethnicities. The 187 participants (99 male and 88 female) were divided into 4 ethnic groups: African, Asian, European, and Hispanic/Latin, with each group containing 29, 60, 68, and 30 samples respectively.

As a prerequisite for participation, it was required that all individuals possess proficient English language skills, which were demonstrated through either their enrollment in English-based academic programs or their utilization of English within their professional settings.

The average age of participants in the study was calculated to be 28.62 years old. The study had originally aimed to

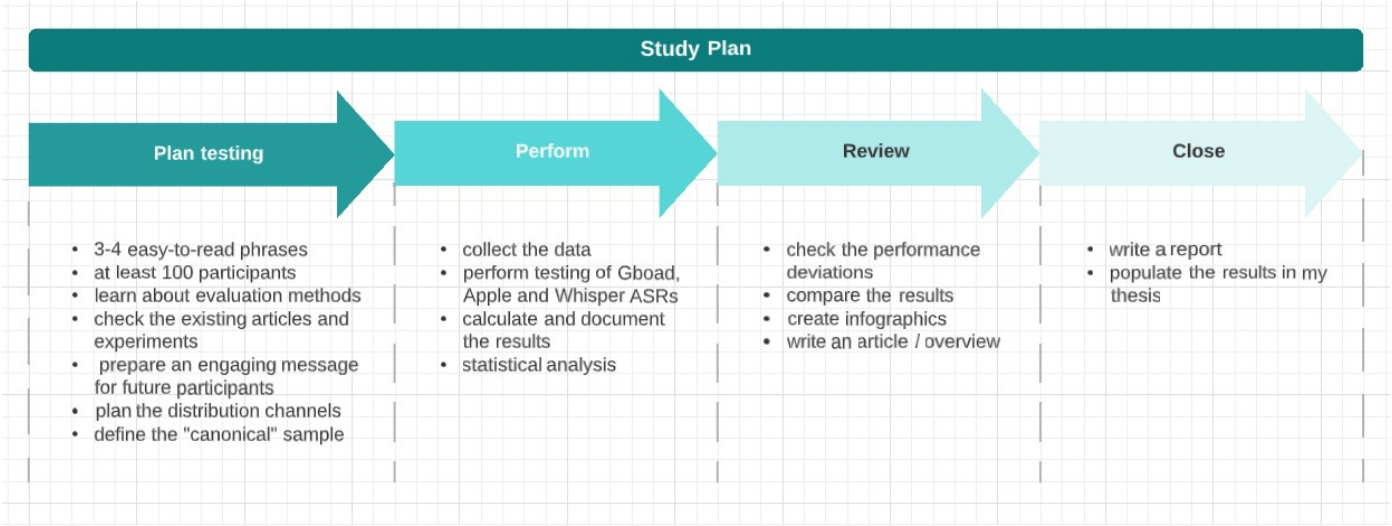


Figure 1: Experiment Planning and Implementation Phases

investigate age-based bias and therefore included the tracking of participants' ages. However, it was found that the majority of participants fell within the age range of 20-30 years. As a result of the limited representation of older individuals, the decision was made to focus on ethnic bias analysis instead.

The following channels were used for the data collection:

- Technical University of Munich student community;
- Tokyo Institute of Technology student community;
- AIESEC organization channels;
- Social Media (LinkedIn, Instagram);
- TUM SEED Center internal channels;
- Friends and acquaintance contacts.

The study was conducted in accordance with ethical considerations. Informed consent was obtained from all participants, and the data was collected and handled in a way to ensure the anonymity and confidentiality of the participants.

5.3. Harvard Sentences

In this study, it was hypothesized that the performance of ASR systems might be influenced by the ethnicity (accent) of a speaker. This was based on the observation that many ASR systems are trained on speech samples from a limited variety of dialects or pronunciation variances (Dahl et al., 2012; Li et al., 2018). In order to investigate this hypothesis, an experimental design was implemented to evaluate the extent to which different accents affected the recognition level of the most commonly used ASR systems. Participants of different ethnicities and accents were asked to speak the same set of phrases from the Harvard sentences.

Harvard sentences are a set of standardized phrases (720 sentences). Developed in the 1950s by Harvard researchers, they are still widely used to test everything from cell phones to Voice over Internet Protocol (VoIP). According to David Pisone, director of the Speech Research Laboratory at Indiana University, Harvard Sentences have become the golden standard for speech-to-text engineers and speech scientists (S. Zhang, 2015).

There are other standardized sets of words and phrases for testing and training ASR, but Harvard sentences are among the oldest and most popular. The sets of sentences are involved in numerous experiments and research in the past, (Schwab et al., 1985) and recent years (Loebach et al., 2010; Smith et al., 2019). The set of phrases within Harvard sentences selected for the study is set H5, sentences 1, 2, and 3 (Harvard Sentences, n.d.).

1. *"A king ruled the state in the early days."*
2. *"The ship was torn apart on the sharp reef."*
3. *"Sickness kept him home the third week*."

The speech samples were analyzed for recognition accuracy and errors. This design allowed for a direct comparison of the ASR systems' performances across different accents, providing insight into the potential influence of accents on recognition level.

5.4. Data Analysis

This sub-chapter presents the methods and results of the data analysis conducted on the results of the experiment, including the implementation of the Word Error Rate (WER) algorithm, Power analysis, Analysis of Variance (ANOVA), and post-hoc Tukey Test.

5.4.1. Word Error Rate Computation

Speech recognition research typically evaluates and compares systems based on the word error rate (WER) metric (Radford et al., 2022). In particular, the performance of ASR systems is measured by computing the WER algorithm for the

transcribed data. Word error rate is a metric that is based on string edit distance. It identifies all differences between the model's output and the canonical sample transcription (Park et al., 2008). The formula for WER is as follows:

$$WER = \frac{S + D + I}{N}$$

where S, D, and I are the number of substitutions, deletions, and insertions respectively, and N is the total number of words in the reference transcription.

In order to minimize human errors and avoid miscalculations, data evaluation was automated. The code was written in JavaScript and used the Google Sheets API to compute the WER for a set of voice transcription data. The key functions of the code are presented and explained below, while the full version of the code is available in *Appendix D. Word Error Rate Implementation* file.

```
function computeErrorRates() {
  const sheet = SpreadsheetApp.
      getActiveSpreadsheet().getSheetByName('Data
      ');
  const canonical = sheet.getRange('K3').
      getValues()[0][0];
  const gboard = sheet.getRange('E3:E192').
      getValues();
  sheet.getRange('D3:D192').setValues(process(
      gboard, canonical));
  const apple = sheet.getRange('G3:G192').
      getValues();
  sheet.getRange('F3:F192').setValues(process(
      apple, canonical));
  const whisper = sheet.getRange('I3:I192').
      getValues();
  sheet.getRange('H3:H192').setValues(process(
      whisper, canonical));
}

function process(entries, expected) {
  return entries.map((entry) => {
    const answer = wer(entry[0], expected);
    return [(answer * 1000 | 0) / 1000];
  });
}

function trim(text, chars) {
  let low = 0;
  for (; low < text.length && chars.includes(text
      [low]); low++);
  let high = text.length - 1;
  for (; high >= low && chars.includes(text[high
      ]); high--);
  return text.slice(low, high + 1);
}

function cleanup(text) {
  return text
  .split(" ")
  .map((s) => trim(s.trim(), ".,_").toLowerCase()
      )
  .filter((s) => s !== "");
}

function wer(text, expected) {
  text = cleanup(text);
  expected = cleanup(expected);
  const n = text.length;
  const m = expected.length;
  const dp = Array.from({ length: n + 1 }, (_, i)
      =>
  Array.from({ length: m + 1 }, (_, j) => (i ===
      0 ? j : j === 0 ? i : 0)) );
  for (let i = 1; i <= n; i++) {
    for (let j = 1; j <= m; j++) {
      dp[i][j] =
      text[i - 1] === expected[j - 1]
      ? dp[i - 1][j - 1]
      : 1 +
      Math.min(Math.min(dp[i - 1][j], dp[i][j -
          1]), dp[i - 1][j - 1]);
    }
  }
  return dp[n][m] / m;
}
```

The code is composed of several functions that work together to calculate the WER.

1. computeErrorRates(): This function is the main function that is called to initiate the computation of word error rates (WER) for the data collected in the experiment. The function starts by getting the active spreadsheet and searching for a sheet named "Data".
2. process(entries, expected): This function takes two inputs, "entries" and "expected", and maps the entries to their corresponding WER values by calling the wer(entry[0], expected) function and returning an array of the WER values truncated to three decimal digits of precision.
3. trim(text, chars): This function is used to remove unwanted characters in the beginning or the end of a string of text. It takes two inputs, "text" and "chars", where "text" is the string of text and "chars" is a string of characters to remove.
4. cleanup(text): This function is used to clean and prepare the text for the WER calculation. It takes one input, "text", and performs several operations such as splitting the text into words, removing unwanted characters and converting all characters to lowercase.
5. wer(text, expected): This function calculates the WER between the given text and the expected text. It takes two inputs, "text" and "expected", and implements a dynamic programming algorithm to find the minimum number of edits required to transform the text into the expected text. The function returns the WER value.

The code first calls the computeErrorRates() function, which initiates the process of computing the WER values for the data. It then uses the getRange() method to get the data from the sheet, and passes it to the process() function along with the canonical text. The process() function then maps the data to their corresponding WER values by calling the wer() function, and sets the values back.

The results for WER computations are presented in the *Appendix A. Voice Transcription Experiment Data* file under the sheet "*Data*".

5.4.2. Power Analysis and ANOVA

In order to determine the statistical significance of the differences in ASR systems' performance between different ethnic groups, we performed Power analysis, ANOVA One-Way with Unequal n's, and post-hoc Honest Significant Difference (HSD) Tukey Test on the results of the experiment.

The power analysis, conducted using the NQuery software, was used to evaluate the effect size and ensure that the experiment had sufficient power to detect meaningful differences between groups (O'Brien & Muller, 1993). The one-way ANOVA, which is a statistical test that compares the means of multiple groups, was conducted using the Excel extension XLMiner Analysis tool pack. This test was used to determine if there were significant differences in ASR performance between the four ethnic groups (African, Asian, European, and Hispanic/Latin) in our study.

In case of significant p-value ANOVA results, we also performed a post-hoc Tukey test in order to compare the pairs of ethnic group performances to each other, and see where the differences lie between the groups (Copenhaver & Holland, 1988; Gleason, 1999; "R: The Studentized Range Distribution", n.d.). Detailed report with the results, tools, and additional files is included in *Appendix B. Anova & NQuery Power & post-hoc Tukey Results* file.

In this study, we accept that the Null Hypothesis is the following: "ASR systems' performance is equally good for all ethnicity groups".

6. Results

In this chapter, we present the results of the experiment examining the performance of ASR systems Gboard, Apple keyboard and Whisper for different ethnic groups. The performance of the ASR systems was measured using the word error rate (WER) algorithm. The results were analyzed using power analysis, one-way ANOVA, and post-hoc Tukey Test.

The results are presented in terms of error rates, tables, and figures, which show the performance of the ASR systems for each ethnic group, as well as the comparison between the ethnic groups.

6.1. Data Analysis Results

The results of the ANOVA One-Way with Unequal n's analysis indicate that there is a statistically significant difference in the performance of the three ASR systems, Gboard, Apple, and Whisper, with respect to recognizing speech from different ethnic groups. The analysis was conducted on 4 groups of participants representing African, Asian, European, and Hispanic/Latin ethnicities. The analysis results can be found in Table 1.

The test significance level, α, which is the probability of rejecting the null hypothesis when it is true, was set at 0.05 for each system. The number of ethnic groups for each tool is 4, the total sample size is 187 participants.

Additionally, the N as multiple of n_1, $\sum r_i$, is 6.448275862 for all systems which indicates that the sample size of each ethnic group is relatively similar for each system which is favorable for comparing the results across the groups and ASR systems.

The variance of means, V, was found to be 0.0034537944 for Gboard, 0.0017757824 for Apple, and 0.0038377534 for Whisper. These results propose that there is a relatively small difference between the means of the scores for the different ethnic groups for ASR Apple, compared to ASR GBoard and ASR Whisper.

The common standard deviation, σ, which is the square root of the variance, was found to be 0.1847353293 for Gboard, 0.2742167753 for Apple, and 0.2175543915 for Whisper. This indicates that the data points are further spread out for Apple than for Gboard and Whisper.

A higher effect size, Δ^2, indicates a larger difference between the groups. This implies that the effect of the ethnic groups on the performance of Apple is lower than the effect of the ethnic groups on the performance of Gboard and Whisper.

Power, the probability of detecting a true effect if one exists, was found to be 96.4% for Gboard, 38.5% for Apple, and 91.4% for Whisper. This suggests that the Gboard and Whisper surveys have a high probability of detecting a true difference if one exists, whereas Apple's power is lower. It is worth noting that the power for Apple is lower than desired ($P_3 \leq 80\%$), which is an indication that the sample size for this system should be increased in future studies.

In order to test the null hypothesis, we perform the One-Way Analysis of Variance (ANOVA) on the tool NQuery for each of the three tools.

The results of the one-way ANOVA for Gboard (Table 2) state that there is a statistically significant difference in the performance of automatic speech recognition systems across different ethnic groups, $F(3, 183) = 3.106$, $p = 0.028$. The null hypothesis, which states that the performance of ASR systems is equally good for all ethnicity groups, is rejected due to the fact that p-value is less than the significance level of 0.05.

The p-value also indicates that there is a 2.8% chance that the observed difference in the performance of Gboard across the ethnic groups is due to random chance, which is considered low and suggests that the difference is likely not due to random variation, but rather a real effect.

As shown in Table 3, the high p-value of 0.481 indicates that there is a 48.1% chance that the observed difference in the performance of Apple keyboard across the ethnic groups is due to random chance, rather than a real effect. This implies that the null hypothesis that the tool (Apple keyboard) has equal performance across ethnic groups cannot be rejected.

The Tukey post-hoc test results for Apple keyboard show that there is no significant difference between the performance of the ASR system for any of the four ethnic groups tested (African, Asian, European, and Hispanic/Latin) (see Table 4). This is indicated by the high p-values for all pairwise comparisons and the inference of "insignificant" for all of them.

Table 1: NQuery One-Way ANOVA with Unequal n's Analysis

	Gboard	***Apple***	***Whisper***
Test Significance Level, α	0.05	0.05	0.05
Number of Groups, G	4	4	4
Variance of Means, V	0.0034537944	0.0017757824	0.0038377534
Common Standard Deviation, σ	0.1847353293	0.2742167753	0.2175543915
Effect Size, $\Delta^2 = V / \sigma^2$	0.101203743	0.0236157482	0.0810851580
Power (%)	**96.39940714**	**38.53636251**	**91.39819872**
N as Multiple of n_1, $\sum r_i = \sum n_i / n_1$	6.448275862	6.448275862	6.448275862
Total Sample Size, N	187	187	187

Table 2: NQuery One-Way ANOVA Results for ASR Gboard

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	*0.3179468*	*3*	*0.10598229*	*3.10551347*	***0.02784065***	*2.65396473*
Within Groups	6.2452669	183	0.03412714			
Total	6.5632138	186				

Table 3: NQuery One-Way ANOVA Results for ASR Apple

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	*0.1861825*	*3*	*0.06206083*	*0.82530756*	***0.48146338***	*2.65396473*
Within Groups	13.761091	183	0.07519721			
Total	13.947273	186				

Table 4: Post-hoc Tukey Results for ASR Apple

treatments pair	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inference
A vs B	0.0099	0.8999947	insignificant
A vs C	0.8398	0.8999947	insignificant
A vs D	1.7755	0.5815922	insignificant
B vs C	1.0389	0.8730357	insignificant
B vs D	2.0577	0.4679167	insignificant
C vs D	1.2597	0.7856598	insignificant

Note: A - African; B - Asian; C - European; D - Latin / Hispanic

The Q statistic, which measures the difference between the means of each pair of groups, is also relatively low for all pairs, further supporting the conclusion of no significant difference in performance across ethnic groups for this ASR system.

The p-value for ASR system Whisper of 0.002 indicates that there is a 0.2% chance that the observed difference in the performance of the ASR system of Whisper across the ethnic groups is due to random chance, which is considered very low and suggests that the difference is likely due to a real effect (Table 5).

The F-value of 5.07 also indicates a significant difference in performance across ethnic groups. This suggests that the null hypothesis that the ASR system Whisper has equal performance across ethnic groups can be rejected.

Overall, the results of the one-way ANOVA test for Whis-

Table 5: NQuery One-Way ANOVA Results for ASR Whisper

Source of Variation	*SS*	*df*	*MS*	*F*	***P-value***	*F crit*
Between Groups	*0.7176599*	*3*	*0.23921996*	*5.06712354*	***0.00214913***	*2.65396473*
Within Groups	8.6394684	183	0.04721021			
Total	9.3571283	186				

per suggest that the null hypothesis can be refuted and the performance of the ASR system likely differs according to ethnic groups.

6.2. Transcription Results

Based on the results presented in Figure 2 and Figure 3, there are significant differences between the performance of ASR algorithms in transcribing the voices of non-native and native English speakers. Non-native speakers had lower transcription rates and higher error rates than native speakers across all ASR algorithms. The highest correct transcription rate was achieved using Whisper by OpenAI (0.866), followed by Gboard by Google (0.778), while the lowest was achieved using Apple keyboard (0.578). On average, the error rate was the highest for Apple keyboard (0.422), followed by Gboard by Google (0.222), while the lowest error rate was achieved using Whisper by OpenAI (0.134).

In contrast, Figure 3 shows that native English speakers had almost perfect transcription rates across all ASR algorithms. The highest transcription rates were achieved using Gboard by Google and Apple keyboard, both scoring a perfect 1.00, while the lowest was achieved using Whisper by OpenAI (0.979) with an insignificant error rate of 0.022.

Overall, the results suggest that ASR algorithms are significantly more accurate at transcribing the speech of native English speakers than non-native English speakers.

When considering the performance of each ASR tool for different ethnic groups, there are also notable differences. Across all ASR tools, Latin/Hispanic speakers had the highest correct transcription rate, while African speakers had the lowest. For Whisper and Gboard, there was a clear trend of increasing performance with Latin/Hispanic and European ethnicities, while for Apple, the performance was relatively consistent across all ethnic groups.

In terms of overall performance, Whisper by OpenAI (see Figure 4) appears to be the most accurate ASR tool with an average correct transcription rate of 86.6% and an average error rate of 13.4%.

Gboard by Google (see Figure 5) also performed relatively well with an average correct transcription rate of 77.8% and an average error rate of 22.2%. However, as shown on Figure 6, Apple keyboard had the lowest performance with an average correct transcription rate of 57.8% and an average error rate of 42.2%.

It is also worth noting that the error rates for Asian and African speakers were consistently lower than those for European and Latin/Hispanic speakers across all ASR tools. Additionally, Whisper by OpenAI appears to have the best performance overall, while the Apple keyboard has the worst performance for non-native speakers.

Overall, the data suggests there are possible biases in the performance of ASR tools, with some ethnic groups experiencing significantly higher error rates than others. It also highlights the need for continued research and development in this area to ensure that AI algorithms are designed and trained in a way that is fair and unbiased for all users.

7. Discussion

The purpose of this thesis was to investigate the hypothesis that the performance of ASR systems may be influenced by the ethnicity (accent) of participants.

The results of the ANOVA analysis show that the null hypothesis is rejected for Gboard and Whisper ASR systems, suggesting that ethnicity has an impact on the performance of these systems. The study showed that there is a statistically significant difference in the performance of ASR systems Gboard, Apple keyboard, and Whisper for participants representing different ethnic groups (African, Asian, European, and Hispanic/Latin).

The results of the ANOVA for the Apple ASR system indicated that the null hypothesis cannot be rejected. However, it is important to note that this system performed exceptionally well for the native English speakers group, implying that there may be issues for non-native English speakers who are not well-represented in the dataset. This highlights the importance of collecting data from a diverse range of participants among non-native English speakers, to ensure that AI algorithms are developed and optimized for a range of users.

The research questions addressed in this thesis were centered on the impact of ethnicity on the performance of ASR systems. The literature analysis suggests that even training and awareness of de-biasing techniques may not be sufficient in eliminating biases in AI algorithms. There are various ways that bias can manifest, and multiple parties are involved in the development and implementation of AI systems.

To further address the issue of bias in AI algorithms, it may be useful to introduce bias review as a mandatory step in the development process. This could involve a team of experts in bias analysis who could provide guidelines or recommendations to avoid possible biases in the design and development of AI algorithms. Such a review process could reduce the burden on individual engineers or developers and make the process more secure. However, further research is

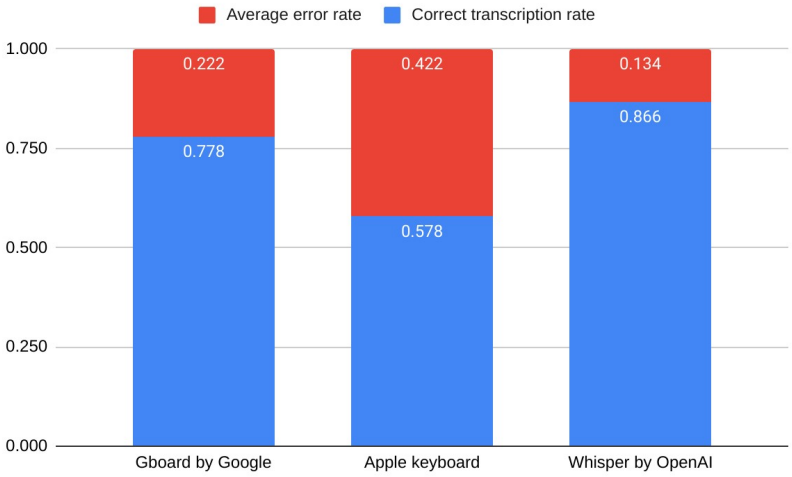


Figure 2: Transcription results for non-native English speakers

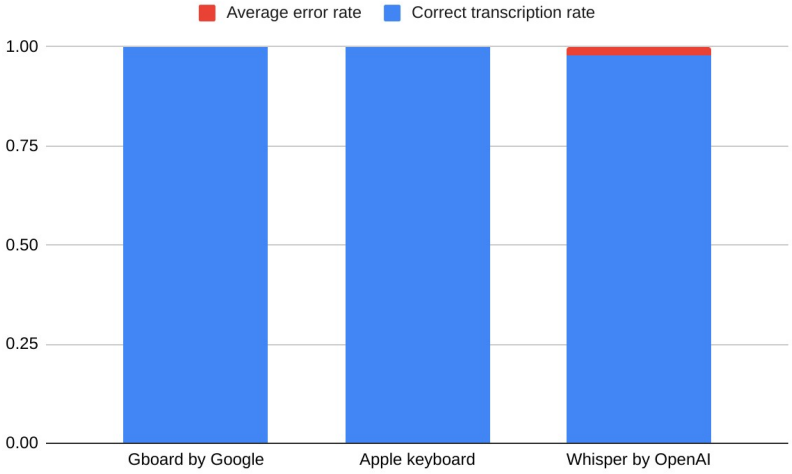


Figure 3: Transcription results for native English speakers

Note: We used canonical Harvard Sentences transcription samples from Open Speech Repository

needed to evaluate the effectiveness of bias review as a tool for reducing biases in AI algorithms.

Additionally, it is crucial to implement strict regulations and conduct thorough checks and inspections before implementing AI technologies.

The findings of this study have important practical implications for the development and use of ASR systems. However, there are certain limitations to this work. The present experiment only focuses on the transcription of audio recordings. It would be more informative to also test the systems on live speech to see if the results generalize to this setting.

As it was discussed previously, not all aspects of the experiment meet the minimum threshold for power, specifically, Apple keyboard, which results in certain findings to be interpreted with caution.

Potentially, further studies will be required with a focus on a more diverse set of phrases in order to evaluate the applications' performance across a wider range of speech patterns and accents, which would provide a more accurate representation of the ASR applications' capabilities. Additionally, it would be more informative for underrepresented groups in the training data sets, where the system may not have seen enough examples to generalize well.

Future research could also explore the specific factors that contribute to the performance differences observed between ASR systems and ethnic groups. This could include exploring the role of language proficiency and exposure. Additionally, further research could investigate ways of improving the accuracy of ASR systems for non-native speakers, such as through accent adaptation or the use of more diverse training data.

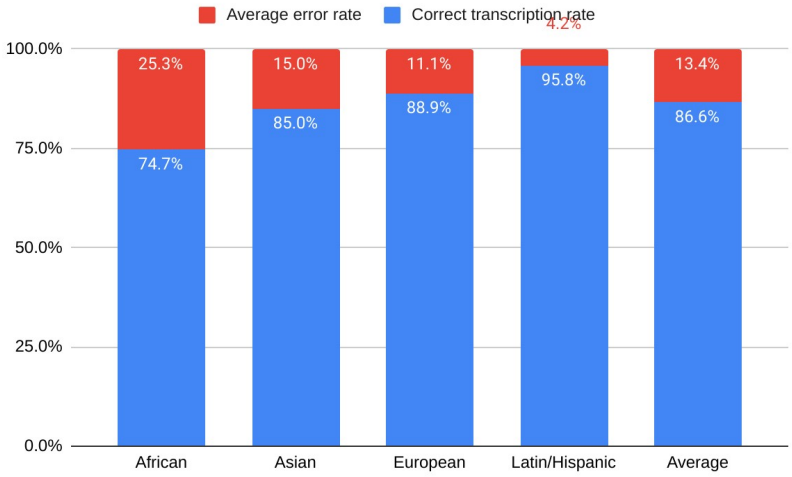


Figure 4: ASR Whisper transcription results

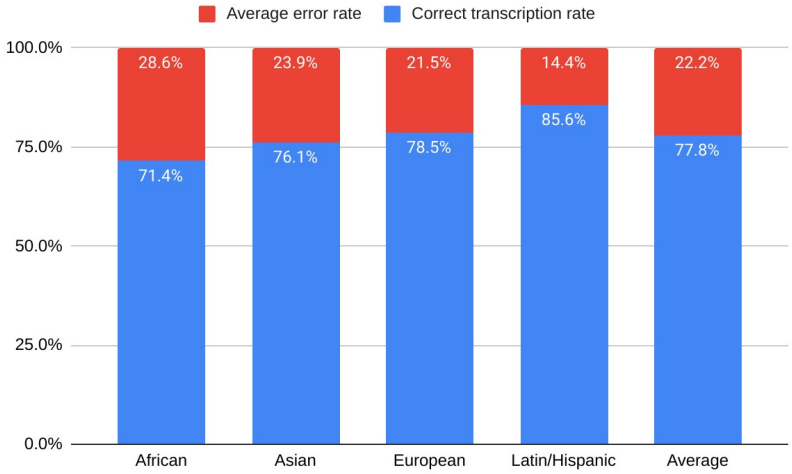


Figure 5: ASR Gboard transcription results

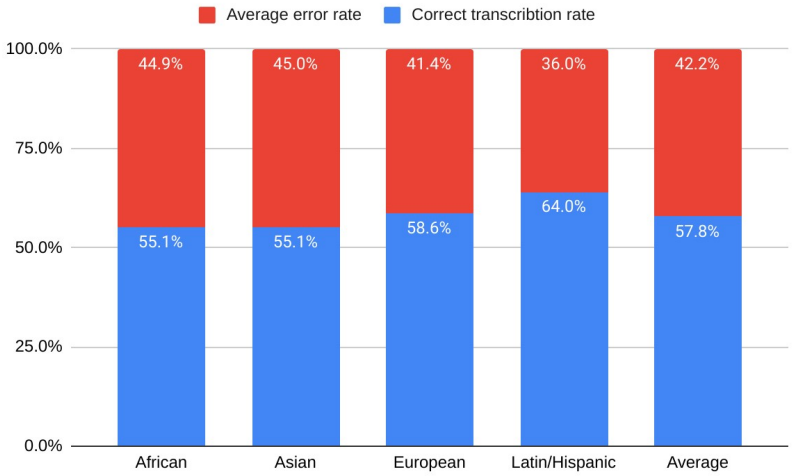


Figure 6: ASR Apple transcription results

Overall, the findings of this thesis suggest that ethnicity (accent) has a significant effect on the performance of ASR systems. It is therefore important for the engineers of AI and ASR systems to be aware of this issue and to take steps to ensure that these systems are not biased toward certain ethnic groups.

8. Conclusion

In conclusion, the results of this thesis demonstrate that ethnicity has a significant impact on the performance of ASR systems. The study provides compelling evidence that ASR systems may not be equally accurate for all users, depending on their ethnicity, which has important practical implications for ASR system development and use. Moving forward, it is important for companies to ensure that their systems are not biased towards certain ethnic groups, and for future research to investigate ways of improving the accuracy of ASR systems for speakers with different accents from all ethnic backgrounds.

While the results of this thesis provide insights into the potential biases of AI algorithms, it is important to note that more data is needed to draw more robust conclusions. Future studies should consider collecting data from a larger and more diverse sample of participants to further investigate the impact of ethnicity and accent on the performance of ASR systems.

References

AARP. (n.d.). AARP Public Policies. Retrieved February 8, 2023, from https://www.aarp.org/about-aarp/policies/?intcmp=FTR-LINKS-INFO-WAS-EWHERE

Age UK & TalkTalk. (2014). Age UK: Helping you love later life - Silversurfers. Retrieved February 7, 2023, from https://www.silversurfers.com/best-of-the-web/lifestyle-best-of-the-web/age-uk-helping-love-later-life/

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., & Rieke, A. (2019). Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. *Proceedings of the ACM on human-computer interaction*, *3*, 199–. http://arxiv.org/pdf/1904.02095

Angwin, J., & Parris, T. (2016). Facebook lets advertisers exclude users by race. Retrieved December 29, 2022, from https://www.cnbc.com/2016/10/28/facebook-lets-advertisers-exclude-users-by-race.html

Angwin, J., Tobin, A., & Varner, M. (2017). Facebook (Still) Letting Housing Advertisers Exclude Users by Race. https://www.propublica.org/article/facebook-advertising-discrimination-housing-race-sex-national-origin

Angwin, J., Larson, J. A., Mattu, S., & Kirchner, L. (2022). Machine Bias. *ProPublica*, 254–264. https://doi.org/10.1201/9781003278290-37

Ansari, T. (2022). OpenAI's Whisper is Revolutionary but (Little) Flawed. Retrieved January 17, 2023, from https://analyticsindiamag.com/openais-whisper-is-revolutionary-but-little-flawed/

App Store. (2016). Gboard – the Google Keyboard. Retrieved January 17, 2022, from https://apps.apple.com/us/app/gboard-the-google-keyboard/id1091700242

Baeza-Yates, R. (2018). Bias on the web. *Communications of the ACM*, *61*(6), 54–61. https://doi.org/10.1145/3209581

Baeza-Yates, R., & Ribeiro-Neto, B. (2011). *Modern Information Retrieval: The Concepts and Technology Behind Search* (2nd ed.). Pearson Education. https://archive.org/details/modern-information-retrieval_202102/page/21/mode/2up

Barocas, S., & Selbst, A. D. (2016). Big Data's Disparate Impact. *California Law Review*, *104*(3), 671–. https://doi.org/10.15779/z38bg31

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. *Neural Information Processing Systems*, *29*, 4356–4364. https://arxiv.org/pdf/1607.06520

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, and Strategies*. Oxford University Press. https://books.google.de/books?id=7_H8AwAAQBAJ&printsec=frontcover#v=onepage&q&f=false

Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Conference on Fairness, Accountability and Transparency*, 77–91.

Cabinet Office Japan. (2017). Annual Report on the Aging Society: 2017 (Summary). https://www8.cao.go.jp/kourei/english/annualreport/2017/2017pdf_e.html

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. https://doi.org/10.1126/science.aal4230

Carvalho, D. T., Pereira, E., & Cardoso, J. S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, *8*(8), 832. https://doi.org/10.3390/electronics8080832

Charles, S. T., & Carstensen, L. L. (2010). Social and Emotional Aging. *Annual Review of Psychology*, *61*(1), 383–409. https://doi.org/10.1146/annurev.psych.093008.100448

Chawla, N. V. (2005). Data Mining for Imbalanced Datasets: An Overview. *Springer eBooks*, 875–886. https://doi.org/10.1007/978-0-387-09823-4_45

Chen, Y., & Petrie, H. (2022). Ageism and Sexism Amongst Young Technicians and Older People in China. *Lecture Notes in Computer Science*, 511–516. https://doi.org/10.1007/978-3-031-08645-8_60

Choi, E. H., Kim, Y. D., Chipalo, E., & Lee, H. C. (2020). Does Perceived Ageism Widen the Digital Divide? And Does It Vary by Gender? *Gerontologist*, *60*(7), 1213–1223. https://doi.org/10.1093/geront/gnaa066

Chu, C., Nyrup, R., Donato-Woodger, S., Leslie, K., Khan, S., Bernett, C., & Grenier, A. (2022). Examining the technology-mediated cycles of injustice that contribute to digital ageism: Advancing the conceptualization of digital ageism: evidence and implications. *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, 545–551. https://doi.org/10.1145/3529190.3534765

Copenhaver, M. D., & Holland, B. (1988). Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. *Journal of Statistical Computation and Simulation*, *30*(1), 1–15. https://doi.org/10.1080/00949658808811082

Czaja, S. J., Lee, C. C., Nair, S. N., & Sharit, J. (2008). Older Adults and Technology Adoption. *Proceedings of the Human Factors and Ergonomics Society ... Annual Meeting*, *52*(2), 139–143. https://doi.org/10.1177/154193120805200201

Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012). Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(1), 30–42. https://doi.org/10.1109/tasl.2011.2134090

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology*, *40*(5), 642–658. https://doi.org/10.1016/j.jesp.2004.02.003

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Retrieved December 20, 2022, from https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

de Paula Couto, M., & Wentura, D. (2017). *Implicit Ageism: Ageism, stereotyping and prejudice against older persons*. The MIT Press. https://doi.org/10.7551/mitpress/10679.003.0006

Dovidio, J. F. (2001). On the Nature of Contemporary Prejudice: The Third Wave. *Journal of Social Issues*, *57*(4), 829–849. https://doi.org/10.1111/0022-4537.00244

Drydakis, N., Macdonald, P. S., Chiotis, V., & Somers, L. (2018). Age discrimination in the UK labour market. Does race moderate ageism? An experimental investigation. *Applied Economics Letters*, *25*(1), 1–4. https://doi.org/10.1080/13504851.2017.1290763

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. S. (2012). Fairness through awareness. *Conference on Innovations in Theoretical Computer Science*. https://doi.org/10.1145/2090236.2090255

EAEA - European Association for the Education of Adults. (2019). New Skills Agenda for Europe - European Association for the Education of Adults. https://eaea.org/our-work/influencing-policy/monitoring-policies/new-skills-agenda-for-europe/

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review*, *109*(3), 573–598. https://doi.org/10.1037/0033-295x.109.3.573

Else-Quest, N. M., & Hyde, J. S. (2016). Intersectionality in Quantitative Psychological Research: Theoretical and epistemological issues. *Psychology of Women Quarterly*, *40*(2), 155–170. https://doi.org/10.1177/0361684316629797

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

Fiske, S., & Taylor, S. (1991). *Social cognition*. Mcgraw-Hill Book Company. https://archive.org/details/socialcognition0002fisk/mode/2up

Floridi, L. (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, *1*(6), 261–262. https://doi.org/10.1038/s42256-019-0055-y

Friemel, T. N. (2016). The digital divide has grown old: Determinants of a digital divide among seniors. *New Media & Society*, *18*(2), 313–331. https://doi.org/10.1177/1461444814538648

Ge, Y., Knittel, C. R., MacKenzie, D., & Zoepf, S. (2016). Racial and Gender Discrimination in Transportation Network Companies. *RePEc: Research Papers in Economics*. https://doi.org/10.3386/w22776

Gleason, J. M. (1999). An accurate, non-iterative approximation for studentized range quantiles. *Computational Statistics & Data Analysis*, *31*(2), 147–158. https://doi.org/10.1016/s0167-9473(99)00002-x

Google. (2017). re:Work - Reduce the influence of unconscious bias with these re:Work tools. Retrieved March 3, 2023, from https://rework.withgoogle.com/blog/fight-unconscious-bias-with-rework-tools/

Google Play. (n.d.). Gboard - the Google Keyboard - Apps on Google Play. Retrieved November 5, 2022, from https://play.google.com/store/apps/details?id=com.google.android.inputmethod.latin

Government of Japan. (2018). Game App Developer in Her 80s Opens ICT World for Fellow Seniors. Retrieved February 15, 2023, from https://www.japan.go.jp/tomodachi/2018/spring-summer2018/game_app_developer.html

Government of Japan. (2019). Restaurant of Mistaken Orders Brings Smiles. Retrieved February 15, 2023, from https://www.japan.go.jp/tomodachi/2019/winter2019/restaurant_of_mistaken_orders.html

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561. https://doi.org/10.1037/pspa0000016

Greenwald, A. G., & Krieger, L. H. (2006). Implicit Bias: Scientific Foundations. *California Law Review*, *94*(4), 945. https://doi.org/10.2307/20439056

Guegan, D., & Hassani, B. K. (2018). Regulatory learning: How to supervise machine learning models? An application to credit scoring. *The Journal of Finance and Data Science*, *4*(3), 157–171. https://doi.org/10.1016/j.jfds.2018.04.001

Guimaraes, A. A. R., & Tofighi, G. (2018). Detecting zones and threat on 3D body in security airports using deep learning machine. *Zenodo (CERN European Organization for Nuclear Research)*. https://doi.org/10.5281/zenodo.1189345

Hajian, S., Bonchi, F., & Castillo, C. F.-D. (2016). Algorithmic Bias. *Knowledge Discovery and Data Mining*, 2215–2216. https://doi.org/10.1145/2939672.2945386

Hao, K. (2019). This is how AI bias really happens — and why it's so hard to fix. Retrieved December 10, 2022, from https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

Harnois, C. E. (2014). Are Perceptions of Discrimination Unidimensional, Oppositional, or Intersectional? Examining the Relationship among Perceived Racial–Ethnic-, Gender-, and Age-Based Discrimination. *Sociological Perspectives*, *57*(4), 470–487. https://doi.org/10.1177/0731121414543028

Harvard Sentences. (n.d.). Harvard Sentences. Retrieved December 5, 2022, from https://harvardsentences.com/#h5-harvard-sentences

Hellström, T., Dignum, V., & Bensch, S. (2020). Bias in machine learning - what is it good for? *European Conference on Artificial Intelligence*, 3–10. http://ceur-ws.org/Vol-2659/hellstrom.pdf

Hobson, S., & Dortch, A. (2022). IBM Policy Lab: Mitigating Bias in Artificial Intelligence. Retrieved February 9, 2023, from https://www.ibm.com/policy/mitigating-ai-bias/

Hultsch, D. F., Hertzog, C., Small, B. J., & Dixon, R. A. (1999). Use it or lose it: Engaged lifestyle as a buffer of cognitive decline in aging? *Psychology and Aging*, *14*(2), 245–263. https://doi.org/10.1037/0882-7974.14.2.245

Hunsaker, A., & Hargittai, E. (2018). A review of Internet use among older adults. *New Media & Society*, *20*(10), 3937–3954. https://doi.org/10.1177/1461444818787348

Keswani, V., Lease, M., & Kenthapadi, K. (2022). Designing Closed Human-in-the-loop Deferral Pipelines. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2202.04718

Kleinberg, J. (2018). Inherent Trade-Offs in Algorithmic Fairness. *Performance evaluation review*, *46*(1), 40. https://doi.org/10.1145/3292040.3219634

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, *10*, 113–174. https://doi.org/10.1093/jla/laz001

Lagacé, M., Charmarkeh, H., Laplante, J., & Tanguay, A. N. (2015). How Ageism Contributes to the Second-Level Digital Divide: The Case of Canadian Seniors. *Journal of technologies and human usability*, *11*(4), 1–13. https://doi.org/10.18848/2381-9227/cgp/v11i04/56439

Landers, R., & Behrend, T. S. (2022). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. *American Psychologist*, *78*(1), 36–49. https://doi.org/10.1037/amp0000972

Lee, N. T. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society*, *16*(3), 252–260. https://doi.org/10.1108/jices-06-2018-0056

Li, B., Sainath, T. N., Sim, K. C., Bacchiani, M., Weinstein, E., Nguyen, P., Chen, Z., Wu, Y., & Rao, K. (2018). Multi-Dialect Speech Recognition with a Single Sequence-to-Sequence Model. *arXiv (Cornell University)*. https://doi.org/10.1109/icassp.2018.8461886

Loebach, J. L., Pisoni, D. B., & Svirsky, M. A. (2010). Effects of semantic context and feedback on perceptual learning of speech processed through an acoustic simulation of a cochlear implant. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(1), 224–234. https://doi.org/10.1037/a0017609

Malisiewicz, T., & Efros, A. A. (2008). Recognition by association via learning per-exemplar distances. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2008.4587462

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, *54*(6), 1–35. https://doi.org/10.1145/3457607

Ministry of Economy, Trade and Industry. (2020). Design Policy Handbook 2020. Retrieved February 15, 2023, from https://www.meti.go.jp/english/press/2020/0420_003.html

Mitzner, T. L., Boron, J. B., Fausset, C. B., Adams, A., Charness, N., Czaja, S. J., Dijkstra, K., Fisk, A. D., Rogers, W. A., & Sharit, J. (2010). Older adults talk technology: Technology usage and attitudes. *Computers in Human Behavior*, *26*(6), 1710–1721. https://doi.org/10.1016/j.chb.2010.06.020

Nellis, S. (2021). Apple sees revenue growth accelerating after setting record for iPhone sales, China strength. Retrieved December 5, 2022, from https://www.reuters.com/article/us-apple-results/apple-tops-wall-street-expectations-on-record-iphone-revenue-china-sales-surge-idUSKBN29W2TD

Niehaves, B., & Plattfaut, R. (2014). Internet adoption by the elderly: employing IS technology acceptance theories for understanding the age-related digital divide. *European Journal of Information Systems*, *23*(6), 708–726. https://doi.org/10.1057/ejis.2013.19

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101–115. https://doi.org/10.1037/1089-2699.6.1.101

O'Brien, R. G., & Muller, K. E. (1993). Unified power analysis for t-tests through multivariate hypothesis. *L. K. Edwards (Ed.), Statistics: Textbooks and monographs, Applied analysis of variance in behavioral science*, *137*, 297–344.

Older Adult Technology Services (OATS). (2022). Digital Equity - OATS. Retrieved February 8, 2023, from https://oats.org/digital-equity/

Olkin, R., & Pledger, C. (2003). Can disability studies and psychology join hands? *American Psychologist*, *58*(4), 296–304. https://doi.org/10.1037/0003-066x.58.4.296

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Broadway Books. https://books.google.de/books?id=NgEwCwAAQBAJ&printsec=frontcover#v=onepage&q&f=false

Open Speech Repository. (n.d.). Open Speech Repository. Retrieved December 6, 2022, from http://www.voiptroubleshooter.com/open_speech/british.html

OpenAI. (2022). OpenAI Charter. Retrieved December 5, 2022, from https://openai.com/charter/

Palmore, E. (2001). The Ageism Survey: First Findings. *Gerontologist*, *41*(5), 572–575. https://doi.org/10.1093/geront/41.5.572

Park, Y., Patwardhan, S. V., Visweswariah, K., & Gates, S. (2008). An empirical analysis of word error rate and keyword error rate. *Conference of the International Speech Communication Association*. https://doi.org/10.21437/interspeech.2008-537

R: The Studentized Range Distribution. (n.d.). http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Tukey.html

Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2212.04356

Schwab, E. C., Nusbaum, H. C., & Pisoni, D. B. (1985). Some Effects of Training on the Perception of Synthetic Speech. *Human Factors*, *27*(4), 395–408. https://doi.org/10.1177/001872088502700404

Shah, D., Schwartz, H. A., & Hovy, D. (2020). Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview. *arXiv (Cornell University)*. https://doi.org/10.18653/v1/2020.acl-main.468

Simon, J. (2022). Amazon SageMaker Clarify Detects Bias and Increases the Transparency of Machine Learning Models | Amazon Web Services. Retrieved February 9, 2023, from https://aws.amazon.com/blogs/aws/new-amazon-sagemaker-clarify-detects-bias-and-increases-the-transparency-of-machine-learning-models/

Smith, G., Pisoni, D. B., & Kronenberger, W. G. (2019). High-Variability Sentence Recognition in Long-Term Cochlear Implant Users. *Ear and Hearing*, *40*(5), 1149–1161. https://doi.org/10.1097/aud.0000000000000691

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising. *HAL (Le Centre pour la Communication Scientifique Directe)*. https://hal.archives-ouvertes.fr/hal-01955343/file/Speicher-etal_PotentialAdDiscrimination_FAT2018.pdf

Srinivasan, R., & Chander, A. (2021). Biases in AI systems. *Communications of The ACM*, *64*(8), 44–49. https://doi.org/10.1145/3464903

Storage, D. (2021). Stereotypes vs. Prejudice vs. Discrimination. https://www.youtube.com/watch?v=6Hr2XpBc_B4

Stypińska, J. (2021). Ageism in AI: new forms of age discrimination in the era of algorithms and artificial intelligence. *Proceedings of the 1st International Conference on AI for People: Towards Sustainable AI, CAIP 2021, 20-24 November 2021, Bologna, Italy*, 39. https://doi.org/10.4108/eai.20-11-2021.2314200

Suresh, H., & Guttag, J. V. (2019). A Framework for Understanding Unintended Consequences of Machine Learning. *arXiv preprint arXiv:1901.10002*, *2*(8).

Torralba, A., & Efros, A. A. (2011). Unbiased look at dataset bias. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2011.5995347

Weiss, R. M., Bass, S. A., Heimovitz, H. K., & Oka, M. (2005). Japan's silver human resource centers and participant well-being. *Journal of Cross-Cultural Gerontology*, *20*(1), 47–66. https://doi.org/10.1007/s10823-005-3797-4

World Health Organization. (2022). Global report on ageism. (License: CC BY-NC-SA 3.0 IGO). Retrieved February 3, 2023, from https://www.who.int/teams/social-determinants-of-health/demographic-change-and-healthy-ageing/combatting-ageism/global-report-on-ageism

Xin, D., Ma, L., Liu, J., Macke, S., Song, S., & Parameswaran, A. (2018). Accelerating Human-in-the-loop Machine Learning. *International Conference on Management of Data*. https://doi.org/10.1145/3209889.3209897

Zafar, M., Valera, I., Rodriguez, M. J., & Gummadi, K. P. (2017). Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment. *The Web Conference*, 1171–1180. https://pure.mpg.de/pubman/item/item_2422970_1/component/file_2422969/arXiv%3A1610.08452.pdf

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B. J., Lyons, T., Manyika, J., Niebles, J. C., Sellitto, M., Shoham, Y., Clark, J. A., & Perrault, C. R. (2021). The AI Index 2021 Annual Report. *arXiv (Cornell University)*. http://arxiv.org/pdf/2103.06312.pdf

Zhang, J., Wu, J., Qiu, Y., Song, A., Li, W., Li, X., & Liu, Y. (2023). Intelligent speech technologies for transcription, disease diagnosis, and medical equipment interactive control in smart hospitals: A review. *Computers in Biology and Medicine*, *153*, 106517.

Zhang, S. (2015). The "Harvard Sentences" Secretly Shaped the Development of Audio Tech. https://gizmodo.com/the-harvard-sentences-secretly-shaped-the-development-1689793568

Zhou, B., Yao, Y., & Luo, J. (2014). Cost-sensitive three-way email spam filtering. *Journal of Intelligent Information Systems*, *42*(1), 19–45. https://doi.org/10.1007/s10844-013-0254-7

Zickuhr, K., & Smith, A. (2012). Older adults and internet use. http://www.pewinternet.org/2012/04/05/older-adults-and-internet-use/

Zimmermann, M. (1986). *Neurophysiology of Sensory Systems*. Springer Nature. https://doi.org/10.1007/978-3-642-82598-9_3

Junior Management Science 9(3) (2024) 1609-1633

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Authentizität und Markenaktivismus – eine empirische Analyse

Authenticity and Brand Activism – An Empirical Analysis

Jaqueline Domnick

Freie Universität Berlin

Abstract

The increasing relevance of brand activism, coupled with its controversial nature, poses challenges for brands. Whether brands should engage in activism and the significance of authentic engagement in this context are questions that need to be addressed. This study employs an experiment to demonstrate that authentic brand activism enhances attitudes towards a brand, with the perceived brand authenticity identified as a key driver. To achieve this, a definition of authentic brand activism is considered, emphasizing the interplay of purpose, values, corporate practices, and activist messaging rather than isolating actions. Furthermore, it is found that the influence of authentic brand activism on perceived brand authenticity is not affected by the moderators of involvement and skepticism. Additionally, preliminary insights into various forms of non-authentic brand activism are gained. This research contributes to the study of the effectiveness of activist engagement and fills an empirical gap regarding the authenticity of brand activism.

Zusammenfassung

Die steigende Relevanz von Markenaktivismus verbunden mit ihrer kontroversen Natur stellt Marken vor Herausforderungen. Ob sich Marken aktivistisch engagieren sollten und wie wichtig dabei das authentische Engagement in diesem Zusammenhang ist, sind Fragen, die es zu beantworten gilt. In dieser Arbeit wird mit Hilfe eines Experiments belegt, dass authentischer Markenaktivismus die Einstellung zu einer Marke verbessert. Dabei wird die dadurch größerer wahrgenommene Markenauthentizität als wichtiger Treiber identifiziert. Dafür wurde eine Definition von authentischem Markenaktivismus betrachtet, bei der nicht die Handlungen isoliert betrachtet werden, sondern das Zusammenspiel von Purpose, Werten, Unternehmenspraktiken und aktivistischer Botschaft. Zusätzlich wird festgestellt, dass der Einfluss von authentischem Markenaktivismus auf die wahrgenommene Markenauthentizität nicht von den Moderatoren Involvement und Skepsis beeinflusst wird. Weiterhin können erste Erkenntnisse über verschiedene Formen von nicht authentischem Markenaktivismus erlangt werden. Diese Arbeit trägt damit zur Forschung über die Wirksamkeit von aktivistischen Engagement bei und schließt eine empirische Lücke in Bezug auf die Authentizität von Markenaktivismus.

Keywords: attitude towards brands; brand activism; brand authenticity

1. Einleitung

In einer globalisierten und digitalisierten Welt, in welcher man eine nahezu uneingeschränkte Auswahl an verschiedenen Marken hat, die jeden erdenklichen Bedarf decken, ist es wichtig, sich von der Konkurrenz abzugrenzen und KundInnen stärker zu binden. Da immer mehr Menschen Marken als Teil ihres Selbstkonzeptes sehen (Aaker, 1997, S. 347), sind

Ich möchte mich herzlich bei Alexandra Rahm bedanken - zusammen haben wir nicht nur die zahlreichen Stunden in der Bibliothek gemeistert, sondern auch in den stressigen Phasen immer wieder etwas zu lachen gefunden. Besonderer Dank gilt meinen Freunden, die sich die Zeit fürs Korrekturlesen und als Sparringpartner genommen haben. Nicht zuletzt danke ich meinem Professor für die hervorragende Betreuung.

DOI: https://doi.org/10.5282/jums/v9i3pp1609-1633

auch die Ansprüche der Konsumierenden an diese größer geworden. Daraus ergeben sich neue Aufgaben und Herausforderungen. So fand die internationale Umfrage der Kommunikationsagentur Edelman heraus, dass etwa ein Drittel der Befragten, die Positionierung und Unterstützung von Marken zu persönlich relevanten Problemen als entscheidenden Faktor sehen, der über den Kauf eines Produktes entscheidet (Edelman, 2019, S. 37). Bereits im Jahr 2016 stufte das Marketing Science Institute die Frage, ob sich Marken gesellschaftspolitisch positionieren sollten, als besonders wichtig für die zukünftige Forschung im Marketing ein (Marketing Science Institute, 2016, S. 17). Es ist daher nicht verwunderlich, dass sich immer mehr Marken zu gesellschaftspolitischen Problemen positionieren, ihren Standpunkt aktiv kommunizieren und sich für ihn einsetzen. Dies wird auch als Markenaktivismus bezeichnet (z.B. Kotler und Sarkar, 2017; Mukherjee und Althuizen, 2020).

Ein Beispiel das zeigt, wie kontrovers Markenaktivismus sein kann, ist die im Jahr 2018 durchgeführte Kampagne von Nike mit dem umstrittenen Footballspieler Colin Kaepernick. Dieser erlangte aufgrund seines Protestes gegen Rassismus, welchen er während seiner Spiele durch einen Kniefall darlegte, große Bekanntheit. Diese Proteste führten vermeintlich zum Ende seiner Karriere bei der National Football League. Während die Kampagne auf der einen Seite als großer Erfolg angesehen wurde und sogar einen Emmy gewann, gab es andererseits drastisches negatives Feedback in Form von Boykotten und dem Verbrennen von Nike-Schuhen (The Guardian, 2019).

Ein weiteres Beispiel, welches zeigt, wie aktuell und umstritten Markenaktivismus sein kann, ist das aktivistische Engagement vieler Marken, ausgelöst durch den Angriffskrieg Russlands auf die Ukraine. Es gibt kaum eine Marke, die sich hierzu nicht geäußert hat. Dies geschieht unter anderem in Form von Statements, der Nutzung der Ukraine-Flagge oder Spenden (z.B. Bücker, 2022). Andere Marken, wie zum Beispiel McDonalds zeigten ihren Standpunkt indem sie drastischen Konsequenzen zogen und alle Filialen in Russland schlossen (z.B. Tagesschau, 2022).

Eine Marke, die in diesem Zusammenhang allerdings mit viel Kritik konfrontiert wurde, ist die Supermarktkette Edeka. Auf Edekas LinkedIn Profil wurde ein Beitrag veröffentlicht, auf dem die ukrainische Flagge mit der Aufschrift „Freiheit ist ein Lebensmittel“, in Verbindung mit einer Solidaritätsbekundung als Bildunterschrift zu sehen ist. Es wurde Edeka vorgeworfen, damit auf geschmacklose Weise Werbung zu machen und den Konflikt für die eigenen Zwecke auszunutzen (EDEKA ZENTRALE Stiftung & Co. KG, 2022a). Weiterhin wurde kritisiert, dass keine Taten folgten und die Solidaritätsbekundung nicht hilfreich sei. Dennoch fand dieser Beitrag auch Zuspruch und das Engagement wurde von einigen positiv gesehen. Kurze Zeit später erschien ein weiterer Beitrag, der mitteilte, dass nun auch Lebensmittel und Drogerieartikel gespendet werden. Einerseits wurde dies gelobt, andererseits wurde weiterhin kritisiert, dass russischer Wodka im Bestand von Edeka verbleibt (EDEKA ZENTRALE Stiftung & Co. KG, 2022b).

Dieses Beispiel verdeutlicht, wie komplex Markenaktivismus ist und greift zusätzlich die unterschiedlichen und sensiblen Reaktionen von Konsumierenden auf vermeintlich altruistische Aktionen auf. Auch die Edelman-Studie weist auf diese Gefahr hin. Während das Positionieren wie zuvor erwähnt, zumeist erwartet und für gut befunden wird, sind 56% derselben Befragten der Meinung, dass das Eintreten für soziale Probleme oft nur als „Marketingtrick“ eingesetzt wird (Edelman, 2019, S. 14). Das Zusammenspiel aus immer größer werdender Relevanz in Verbindung mit der kontroversen und sensiblen Natur des Markenaktivismus macht das Thema zu einem interessanten und praxisrelevanten Untersuchungsobjekt.

Insgesamt lässt sich festhalten, dass Markenaktivismus ein Thema ist, das sehr kritisch angesehen wird und viele konträre Meinungen hervorruft. Aus diesem Grund beschäftigt sich die vorhandene Literatur hauptsächlich mit der Wirksamkeit von Markenaktivismus unter verschiedenen Bedingungen. Ein zentrales Ergebnis in diesem Zusammenhang ist, dass vor allem die Übereinstimmung von Konsumierenden mit dem Standpunkt der Marke über Erfolg oder Misserfolg entscheidet (Hydock et al., 2020; Mukherjee & Althuizen, 2020). Ein weiterer Aspekt, der oft behandelt wird, ist die Authentizität von Markenaktivismus (z.B. Mirzaei et al., 2022; Vredenburg et al., 2020), die quantitativ bisher lediglich von Hydock et al. (2020) als Moderator untersucht wurde.

Die grundlegende Bedeutung der Authentizität kann im Markenkontext nicht ignoriert werden. Authentizität ist eine Kernkomponente nachhaltig erfolgreicher Marken, da sie Teil der einzigartigen Markenidentität ist (Beverland, 2005, S. 1003 & 1025). Zudem ist Authentizität wichtig für das Selbstkonzept eines Menschen, da sie zielrelevant ist und daher dringend angestrebt wird (Napoli et al., 2014, S. 1090).

Somit stellt fehlende Authentizität von Markenaktivismus ein großes Problem für die Marke dar. Häufig werden, wie das Beispiel von Edeka und die Umfrage von Edelman zeigen, die Motive hinter dem Markenaktivismus angezweifelt und er wird als irreführend angesehen. Dies kann dazu führen, dass im Gegensatz zu ehrlichem Interesse, opportunistische Gedanken hinter dem aktivistischen Engagement vermutet werden. Dieses Phänomen, das auch als „Woke Washing“ bezeichnet wird, kann schließlich zu einer negativen Einstellung gegenüber der Marke und den daraus resultierenden Konsequenzen führen (Vredenburg et al., 2020, S. 449). Gerade bei Themen, die einen hohen gesellschaftlichen Druck ausüben, wie z.B. der Krieg in der Ukraine, sollte immer die Frage gestellt werden, ob eine Marke wirklich uneigennützig und wertegetrieben handelt oder ob sie sich dadurch Vorteile erhofft und gesellschaftliche Sanktionen vermeiden möchte.

Durch die Komplexität und Vielschichtigkeit des Authentizitätskonstruktes sowie dem Mangel an empirischer Forschung wird in dieser Arbeit speziell der Einfluss von Authentizität auf Markenaktivismus empirisch untersucht. Die Frage, die primär beantwortet werden soll, ist, welchen Einfluss authentischer bzw. nicht authentischer Markenaktivismus auf die Einstellung gegenüber der Marke hat. Das Inter-

esse an der Einstellung besteht vor allem dadurch, das sich andere Phänomen wie die Markenloyalität oder die Kaufentscheidung daraus ableiten lassen (z.B. Baldinger und Rubinson, 1996; Spears und Singh, 2004). Um erklären zu können, was als authentischer bzw. nicht authentischer Markenaktivismus bezeichnet werden kann, wird in der vorliegenden Arbeit der Definition von Vredenburg et al. (2020) gefolgt. So ist Markenaktivismus nur dann authentisch, wenn der Purpose der Marke, ihre Werte, ihre aktivistische Botschaft und ihre pro-sozialen Unternehmenspraktiken zueinander passen (Vredenburg et al., 2020, S. 450). Wenn die aktivistische Botschaft beispielsweise nicht von den Werten, dem Purpose und den Unternehmenspraktiken gestützt wird, kann der Markenaktivismus hingegen nicht als authentisch angesehen werden (Vredenburg et al., 2020, S. 452). Zum Beispiel wurde so das Engagement der Marke Gilette für die „Me Too"-Bewegung als ein Paradebeispiel für Woke Washing angesehen, da sich die Marke einerseits für die Rechte der Frauen einsetzte, andererseits aber über viele Jahre hinweg toxische männliche Werte predigte (Iqbal, 2019). Diese Unstimmigkeit führt dazu, dass Konsumierende das Engagement als nicht authentisch wahrnehmen.

In anderen Forschungsarbeiten fehlt diese ganzheitliche Betrachtung der Authentizität des Markenaktivismus. Am Beispiel der Forschung von Hydock et al. (2020) wird nur der Markenaktivismus als Ganzes als authentisch oder nicht authentisch betrachtet, nicht aber, wie er zustande kam. Wiederum in anderen Untersuchungen wird die Passung zwischen dem Thema des Markenaktivismus und dem Geschäftsfeld der Marke für die Bewertung der Authentizität herangezogen (z.B. Champlin et al., 2019; Mirzaei et al., 2022; Nan und Heo, 2007).

Ein weiterer Aspekt, der beachtet werden muss, ist, dass die Wahrnehmung von Authentizität subjektiv sein kann (Grayson & Martinec, 2004, S. 298). Es wird daher weiter untersucht, ob der definierte authentische Markenaktivismus tatsächlich authentisch erscheint und ob die Wahrnehmung durch weitere Faktoren beeinflusst werden kann. Hierfür werden die Moderatoren Skepsis und Involvement herangezogen. Skepsis wurde als Moderator gewählt, da aus der bestehenden Literatur zu erwarten ist, dass hier ein Zusammenhang besteht (z.B. Morhart et al., 2015; Skarmeas und Leonidou, 2013). Involvement ist interessant zu betrachten, da behandelte Themen häufig gesellschaftlich relevant, aber unterschiedlich emotional behaftet sind.

Zusammenfassend ergeben sich folgende Forschungsfragen, die mit Hilfe eines Experiments beantwortet werden sollen:

> *F1) „Wie wirkt sich authentischer (vs. nicht authentischer) Markenaktivismus auf die Einstellung zur Marke aus?"*
>
> *F2) „Kann die Konzeptualisierung von authentischem und nicht authentischem Markenaktivismus nach Vredenburg et al. (2020) empirisch belegt werden?"*
>
> *F3) „Lässt sich die Wahrnehmung von authentischem (vs. nicht authentischem) Markenaktivismus durch die Variablen Skepsis oder Involvement beeinflussen?"*

Hierfür werden zunächst die theoretischen Grundlagen der im konzeptuellen Modell verwendeten Variablen erläutert. Es wird zuerst das allgemeine Thema Markenaktivismus thematisiert und insbesondere die unabhängige Variable authentischer (vs. nicht authentischer) Markenaktivismus definiert. Anschließend wird das Konzept der wahrgenommenen Markenauthentizität vorgestellt, welches im Modell als Mediator fungiert. Danach wird kurz theoretisch erläutert, weshalb die Einstellung gegenüber der Marke ein interessante abhängige Variable darstellt. Zum Schluss werden die theoretischen Grundlagen der beiden Moderatorvariablen Skepsis und Involvement wiedergebeben. Nachdem die Variablen einzeln vorgestellt wurden, werden sie im nächsten Schritt zueinander in Beziehung gesetzt. Die in dieser Arbeit untersuchten Hypothesen werden hergeleitet. Daraufhin folgt eine Übersicht des methodischen Vorgehens. Hier wird näher auf die Durchführung des Experiments, die Operationalisierung der Variablen und die Auswertung der Ergebnisse eingegangen. Die gewonnenen Ergebnisse werden im letzten Abschnitt ausführlich diskutiert. Die Diskussion beinhaltet eine Zusammenfassung und Einordnung der Ergebnisse, Implikationen für Theorie und Praxis sowie Limitationen und Vorschläge für zukünftige Forschung.

2. Theoretische Grundlage

2.1. Markenaktivismus

2.1.1. Einführung und Abgrenzung

Unter Markenaktivismus wird allgemein verstanden, dass Marken ihre eigenen Ansichten zu wirtschaftlichen, sozialen, ökologischen oder politischen Problemen teilen, mit dem Ziel diese positiv zu beeinflussen und Veränderungen herbeizuführen (Kotler & Sarkar, 2017). Die Umsetzung kann dabei sehr vielseitig sein und es unterliegt keiner strengen Definition, was unter Markenaktivismus zu verstehen ist und was nicht (z.B. Mirzaei et al., 2022; Mukherjee und Althuizen, 2020; Shetty et al., 2019). Dieser kann sich, wie am Beispiel von Nike (The Guardian, 2019), in Form von Marketingkampagnen oder wie am Beispiel von Edeka (EDEKA ZENTRALE Stiftung & Co. KG, 2022a, 2022b), durch Statements und Handlungen zeigen. Markenaktivismus kann sich auch in Form von organisatorischen Maßnahmen zeigen, wie z.B. der Entscheidung bestimmte Orte, Personen oder Institutionen zu boykottieren, wie das bereits dargestellte Beispiel von McDonalds verdeutlicht (Tagesschau, 2022). Dass der Kreativität bei der Gestaltung keine Grenzen gesetzt werden, zeigt die Marke Ben & Jerry's, die den Namen ihrer Eiscremesorte „Hubby Chubby" in „Hubby Hubby" änderte, um sich damit für die gleichgeschlechtliche Ehe einzusetzen (Huffington Post, 2011).

Auch wenn Markenaktivismus in der Form und damit auch in der Intensität sehr unterschiedlich sein kann, sollte das Ziel der positiven Beeinflussung des öffentlichen Problems im Fokus stehen (Kotler und Sarkar, 2017; Vredenburg et al., 2020, S. 445). Häufig wird in diesem Kontext auch von Corporate Political Advocacy oder Corporate Social Advocacy gesprochen. Diese Begriffe können sinngemäß zum Überbegriff Markenaktivismus zugeordnet werden (Hydock et al., 2020, S. 1136). Deutlich abzugrenzen ist Markenaktivismus jedoch von dem verwandten Konzept Corporate Social Responsibility (CSR). Anhand des Vergleichs mit CSR werden auch einige Charakteristika von Markenaktivismus deutlich.

Der Begriff CSR wird von der EU-Kommission definiert als „die Verantwortung von Unternehmen für ihre Auswirkungen auf die Gesellschaft“ (EU-Kommission, 2011, S. 7). Aus dieser Definition geht hervor, dass sich die Verantwortung der Unternehmen vor allem aus den eigenen wirtschaftlichen Aktivitäten ergibt (Wettstein & Baur, 2016, S. 204). Am Beispiel einer Bekleidungsfirma entsteht somit unter anderem die Verantwortung für Unternehmen, die ArbeiterInnen fair zu bezahlen und Umweltschäden zu minimieren. Häufige Themen, die von CSR abgedeckt werden, sind Umweltschutz, Katastrophenhilfe, Kinderarbeit, Engagement für die Gemeinschaft, fairer Handel, Armut, Hunger und Bildung (Peloza & Shang, 2011, S. 122ff.). Ähnlich wie beim Markenaktivismus kann sich CSR vor allem in unternehmerischen Maßnahmen manifestieren, wie z.B. der Einbindung nachhaltiger Unternehmenspraktiken oder der Zahlung fairer Löhne an Arbeitnehmende, insbesondere in Entwicklungsländern (z.B. Collins et al., 2007). Andererseits liegt ein großer Schwerpunkt der CSR-Umsetzung auf freiwilligen Hilfsprogrammen und Spenden (Peloza & Shang, 2011, S. 120). Im Gegensatz zu CSR Aktivitäten, die für gewöhnlich in der Unternehmensstrategie eingebettet sind, erfolgt das Engagement in Markenaktivismus meist spontan (Mukherjee & Althuizen, 2020, S. 774). Markenaktivismus ergibt sich in den meisten Fällen aus aktuellen gesellschaftlich relevanten Problemen, wie z.B. den „Black Lives Matter“-Protesten oder der „Me Too“-Bewegung (Mirzaei et al., 2022, S. 1). Die Verantwortung wird hier eher auf einer Werteebene gesehen, sodass die Marke eine Verpflichtung verspürt, sich mit dem Problem auseinanderzusetzen und entsprechend zu positionieren (Vredenburg et al., 2020, S. 449).

Der entscheidendste Aspekt, in dem sich CSR und Markenaktivismus unterscheiden, ist, wie sie wahrgenommen werden. 2017 war die öffentliche Meinung zu der Frage, ob die USA die Verantwortung haben, syrische Flüchtlinge aufzunehmen, fast gleichmäßig geteilt – 47% der Amerikaner waren dafür, 49% dagegen (Pew Research Centre, 2017, S. 21). Das klare Positionierung der Marke Starbucks, syrische Flüchtlinge aufzunehmen und diese zusätzlich zu beschäftigen, führte dementsprechend sowohl zu positiven als auch zu negativen Reaktionen (Reuters, 2017). Ein weiteres Beispiel, welches unterschiedliche Reaktionen bei der Gesellschaft auslöste ist das der amerikanischen Fast-Food-Kette Chick-fil-A, die sich 2012 für ein Verbot der gleichgeschlechtlichen Ehe einsetzte. So haben Befürwortende einen „Chick-Fil-A Appreciation Day“ ins Leben gerufen, der zu einer 30-prozentigen Umsatzsteigerung im Vergleich zu einem normalen Tag geführt hat. Auf der anderen Seite hat dieses Engagement zu starken Protesten gegen die Marke, vor allem seitens LGBTQ+ AktivistInnen geführt (Norman, 2012). Es gibt zu vielen Problemen keinen Konsens in der Gesellschaft, weshalb auch die Reaktionen gespalten sind und die allgemeine Wirksamkeit von Markenaktivismus eher kritisch gesehen wird (Hydock et al., 2020, S. 1136; Mukherjee und Althuizen, 2020, S. 3; Vredenburg et al., 2020, S. 448). CSR hingegen wird in der Gesellschaft überwiegend sehr positiv wahrgenommen und ist schon lange eine gängige Praxis, um die Unternehmensreputation zu erhöhen (z.B. Chernev und Blair, 2015; Hydock et al., 2020, S. 1136).

2.1.2. Bisherige Forschung

Aufgrund der kontroversen Natur und den unklaren Auswirkungen von Markenaktivismus beschäftigt sich die bestehende Literatur vorwiegend damit, ob er positiv oder negativ für Marken ist. So kam McDonnell (2016, S. 66f.) zu dem Ergebnis, dass eine Extremform des Markenaktivismus, der von einem Unternehmen gesponserte Boykott, eine Taktik ist, die nur dann genutzt werden sollte, wenn das Unternehmen keine Alternative sieht und selbst an Reputationsdefiziten leidet.

Ein Kernelement, das über Erfolg oder Misserfolg von Markenaktivismus entscheidet, ist die Zustimmung zu dem von der Marke vertretenen Standpunkt (z.B. Hydock et al., 2020; Mukherjee und Althuizen, 2020). Dabei werden Auswirkungen vor allem anhand der finanziellen Leistung, der Kaufabsicht oder der Einstellung gegenüber der Marke gemessen (z.B. Dodd und Supa, 2015; Dodd und Supa, 2014; Mukherjee und Althuizen, 2020). Bezüglich der Übereinstimmung des Standpunktes, wird von Mukherjee und Althuizen (2020, S. 783f.) auf die asymmetrische Natur des Markenaktivismus aufmerksam gemacht. Demnach hat die Meinungsübereinstimmung keinen positiven Effekt. Lediglich Meinungsverschiedenheiten haben einen negativen Effekt. In diesem Zusammenhang wird von Hydock et al. (2020, S. 1149) die Unternehmensgröße als entscheidender Faktor für den Erfolg der Implementierung von Markenaktivismus identifiziert. Für ein Unternehmen mit hohem Marktanteil besteht eine größere Gefahr, viele bestehenden KundInnen zu verlieren, die nicht mit dem eigenen Standpunkt übereinstimmen. Verglichen mit dem potentiellen Verlust, besteht eine geringere Chance neue KundInnen hinzuzuzugewinnen. Kleinere Unternehmen, vor allem die, die Nischenmärkte bedienen, können hingegen davon profitieren, dass sie im Vergleich zu großen Unternehmen weniger KundInnen verlieren und im Gegensatz dazu, mehr neue gleichgesinnte Menschen hinzugewinnen.

Weiterhin kann der Erfolg auch von der Zielgruppe abhängen, an die sich der Markenaktivismus richtet. So fanden Shetty et al. (2019, S. 164) heraus, dass Millennials positiver auf Markenaktivismus reagieren, da diese Generation als besonders aufgeklärt gilt und ihre Verantwortung gegenüber der Gesellschaft und der Erde sieht.

Weitere Forschung beschäftigt sich auch damit, welche Rolle Authentizität im Bereich Markenaktivismus einnimmt. Zum Beispiel untersucht Sobande (2019) näher das Phänomen Woke Washing. Empirische Forschung in Form eines Experiments wurde nur von Hydock et al. (2020, S. 1148f.) durchgeführt, mit dem Ergebnis, dass niedrige Authentizität zwar die positiven Effekte von Markenaktivismus mindert, nicht aber die negativen Effekte. Mirzaei et al. (2022, S. 10) ermitteln durch eine Inhaltsanalyse sechs Dimensionen, die bei der Erreichung von authentischem Markenaktivismus berücksichtigt werden müssen. So muss beachtet werden wie unabhängig das Thema von sozialen Trends ist, wie neutral die Botschaft ist, wie sehr die Marke hält, was sie verspricht, welche Motivation sie verfolgt und zuletzt wie sehr das Thema zum Geschäft der Marke passt. Letzteres wird auch in den Forschungen von Champlin et al. (2019), Key et al. (2021) und Vredenburg et al. (2020) aufgenommen.

Vredenburg et al. (2020) definieren ebenfalls Faktoren, die zu authentischem Markenaktivismus führen. Diese werden im nächsten Kapitel näher erläutert und bilden die Grundlage für die Definition von authentischem Markenaktivismus in dieser Arbeit.

2.1.3. Authentischer Markenaktivismus

Wenn Markenaktivismus nicht authentisch ist, wird kein ehrliches Engagement dahinter vermutet, was zu Misserfolgen führen kann. Schlimmer noch als das positive Effekte ausbleiben kann dies dazu führen, dass sich KundInnen aktiv hintergangen fühlen (Forehand und Grier, 2003, S. 350; Vredenburg et al., 2020, S. 451). Daher sollte die Gefahr von nicht authentischem Markenaktivismus nicht unterschätzt werden. Dies macht eine Abgrenzung dessen, was als authentischer und was als nicht authentischer Markenaktivismus gilt, notwendig.

Das Problem der nicht authentischen Darstellung ist ebenso beim CSR-Konzept in Form von Greenwashing bekannt. Beim Greenwashing werden Unternehmen fälschlicherweise als besonders umweltfreundlich und verantwortungsvoll dargestellt. Dabei wird das umweltfreundliche Verhalten aktiv betont und das umweltschädlich Verhalten verschleiert. Zu diesem Zweck werden oft öffentlichkeitswirksame Kampagnen für umweltfreundliches Engagement initiiert, um von den tatsächlichen umwelt- und sozialschädlichen Unternehmenspraktiken abzulenken (Marquis et al., 2016, S. 483). So versucht beispielsweise der Fast-Fashion-Riese Primark mit dem Konzept „Primark Cares“ Werbung für nachhaltige Mode zu machen und ein besseres Image zu bekommen, während die Produktion dieser nachhaltigeren Mode aber immer noch unter äußerst schlechten Bedingungen stattfindet und weiterhin der Großteil der Produkte nicht nachhaltig ist (Changing Markets Foundation, 2022).

Analog dazu wird der unehrliche Versuch, sich aktivistisch für ein gesellschaftspolitisches Thema einzusetzen als Woke Washing bezeichnet. Genauer verstehen insbesondere Vredenburg et al. (2020, S.449) darunter, dass sich eine Marke zwar für soziale Gerechtigkeit einsetzt, aber weder in ihrer grundlegenden Mission noch in ihren Werten oder Unternehmenspraktiken eine Ausrichtung auf soziales Verhalten zeigt und somit kein ernsthaftes Interesse an der Lösung des Problems besteht. Demnach kann Woke Washing als Gegenpart zu authentischem Markenaktivismus gesehen werden. Vredenburg et al. (2020) definieren genauer, welche Komponenten beachtet werden müssen um authentischen von nicht authentischem Markenaktivismus zu unterscheiden. Sie definieren authentischen Markenaktivismus als „purpose- and values-driven strategy in which a brand adopts a nonneutral stance on institutionally contested sociopolitical issues, to create social change and marketing success“ (Vredenburg et al., 2020, S. 446). Diese Definition ergänzt die eingangs genannte allgemeine Definition von Kotler und Sarkar. Es wird betont, dass Markenaktivismus vom sogenannten „Purpose“ angetrieben werden sollte. Darunter versteht man, den übergeordneten Zweck oder die Mission die eine Marke verfolgt, die über das eigene wirtschaftliche Interesse hinausgeht (Accenture, 2019, S. 6f.). Außerdem geht aus der Definition genauer hervor, dass es sich um eine nicht neutrale Meinung zu einem umstrittenen Thema handelt und, dass das Engagieren in authentischem Markenaktivismus auch zu Marketingerfolg führt. Besonders der letzte Punkt, der Marketingerfolg, sollte kritisch betrachtet werden. Wie die Forschung zeigt, führt Markenaktivismus in vielen Fällen nicht zu diesem Erfolg (z.B. Mukherjee und Althuizen, 2020). Außerdem würde eine Fokussierung auf Marketingerfolg den eigentlichen sozialen Hintergedanken überlagern (Skarmeas & Leonidou, 2013, S. 1833f.) und dazu führen, dass der Markenaktivismus eher als nicht authentisch wahrgenommen wird (Vredenburg et al., 2020, S. 451).

So benennen Vredenburg et al. (2020) vier Faktoren, deren gemeinsames Zusammenspiel zu authentischem Markenaktivismus führt – die Botschaft, die pro-sozialen Unternehmenspraktiken, den Purpose und die Werte der Marke. Diese werden nachfolgend kurz erläutert.

Botschaft

Was sowohl authentischer als auch nicht authentischer Markenaktivismus gemeinsam haben, ist das Kommunizieren einer aktivistischen Botschaft zu einem kontroversen Thema. Es wird klar Stellung zu einem gesellschaftlichen Problem bezogen, sodass die Position der Marke deutlich wird (Vredenburg et al., 2020, S. 446). Die Form und das Ausmaß können sich dabei, wie in Kapitel 2.1.1 beschrieben unterscheiden.

Pro-Soziale Unternehmenspraktiken

Nur eine Marke, die authentischen Markenaktivismus verfolgt, wird diese Botschaft auch mit entsprechenden Taten stützen. Da das Ziel von Markenaktivismus ist, Veränderung in der Zukunft zu bewirken, werden auch die Unternehmenspraktiken primär nach ihrer Zukunftsorientierung bewertet. So wird eine einmalige Spende weniger positiv bewertet als eine dauerhafte Veränderung der Unternehmenspraktiken oder eine kontinuierliche Unterstützung für

einen bestimmten Zweck (Vredenburg et al., 2020, S. 448). Wird die Botschaft nicht durch entsprechende pro-soziale Unternehmenspraktiken untermauert oder wird deren Fehlen aktiv verschleiert, können keine aufrichtigen Absichten hinter dem Engagement erkannt werden (Vredenburg et al., 2020, S. 451)

Es ist wichtig, kritisch gegenüber den Handlungen zu sein, da KundInnen aufgrund von Informationsasymmetrien nur ein bedingtes Wissen über eine Marke und ihre unternehmerischen Aktivitäten haben (Mishra et al., 1998, S. 277), wodurch es zu Täuschungen kommen könnte. Zum Beispiel ist das Beenden aller Geschäftsbeziehungen zu Russland auf Grund des Ukraine Krieges, nur dann wirksam, wenn dies auch konsequent durchgesetzt wird. Wird beispielsweise durch Subunternehmer weiterhin indirekt Handel betrieben, könnte die vermeintlich gute Tat als Woke Washing angesehen werden.

Pro-Soziale Werte und Purpose

Genauso wie pro-soziale Unternehmenspraktiken vorhanden sein müssen, um die Botschaft zu unterstützen, ist es ebenso notwendig, dass auch pro-soziale Werte vertreten werden und die Marke einen Purpose hat, der über das Unternehmensinteresse hinausgeht (Vredenburg et al., 2020, S. 450). Wenn dies nicht der Fall ist, werden hinter dem Markenaktivismus keine ehrlichen Absichten vermutet.

Auch wenn Vredenburg et al. (2020) explizit von prosozialen Werten sprechen, können auch regressive Standpunkte authentisch sein, sofern Marken in ihrem Wertesystem pro- sozial handeln (Vredenburg et al., 2020, S. 451). Aus diesem Grund kann auch Markenaktivismus gegen die gleichgeschlechtliche Ehe authentisch sein, wenn die Marke für traditionelle Werte einsteht. Auf der anderen Seite würde eine Marke, die sich öffentlich für Frauenrechte einsetzt, gleichzeitig aber eine gender-pay gap hat und Frauen in den eigenen Reihen diskriminiert, nicht authentisch angesehen werden. Die Unterstützung der Frauenrechte würde hier nicht zu den eigenen gelebten Werten passen.

Daraus ergibt sich, dass Markenaktivismus nur dann authentisch sein kann, wenn eine Passung des Purpose der Marke, ihrer Werte, ihrer aktivistische Botschaft und ihren prosozialen Unternehmenspraktiken vorliegt und die Hauptintention ist, eine Lösung für ein sozio-politisches Problem zu finden. Ist hier keine Passung vorhanden, wird der

Markenaktivismus als nicht authentisch angesehen, da höchstwahrscheinlich nur opportunistisch gehandelt wird und das Engagement lediglich zum eigenen Vorteil getätigt wird (Vredenburg et al., 2020, S. 449f.).

2.2. Wahrgenommene Markenauthentizität

Wie KundInnen eine Marke bewerten, hängt unter anderem davon ab, wie authentisch eine Marke ist (z.B. Fritz et al., 2017; Morhart et al., 2015; Napoli et al., 2014). Da Authentizität jedoch kein rein objektives Konstrukt ist, ist es notwendig, die wahrgenommene Markenauthentizität näher zu betrachten (Grayson & Martinec, 2004, S. 298). Um die Notwendigkeit dieser Spezifikation zu verstehen und das Authentizitätskonstrukt allgemein besser greifen zu können, sollten drei philosophische Perspektiven genauer betrachtet werden (Morhart et al., 2015, S. 201).

Aus objektivistischer Sicht hat die Authentizität ihren Ursprung in Museen, wo Experten und Expertinnen die Echtheit von Kunstobjekten prüfen und das Original von seinen Kopien unterscheiden (Trilling, 1972, S. 93). Nach dieser Auffassung beruht die Authentizität auf einer evidenzbasierten Realität, die anhand äußerer Merkmale bestimmt werden kann (Grayson & Martinec, 2004; Trilling, 1972). Grayson und Martinec (2004, S. 298) bezeichnen diese Authentizität als indexikalisch. Für den Markenkontext bedeutet das, dass Authentizität anhand überprüfbarer Informationen, wie z.B. am Alter oder an den Inhaltsstoffen bewertet werden kann (Morhart et al., 2015, S. 201).

Aus einer konstruktivistischen Perspektive ist Authentizität ein gesellschaftlich erschaffenes Konstrukt. So gibt es nicht die eine Realität, anhand derer Authentizität bewertet wird, sondern vielmehr verschiedene Interpretationen davon (Grayson & Martinec, 2004, S. 298). Authentizität wird nicht als eine einem Objekt innewohnende Qualität betrachtet, sondern als eine Projektion der eigenen Überzeugungen, Erwartungen und Perspektiven (Wang, 1999, S. 351). Diese Art von Authentizität wird auch als ikonisch bezeichnet (Grayson & Martinec, 2004, S. 298). Für den Markenkontext bedeutet das, dass im Gegensatz zu den objektiven Eigenschaften einer Marke, vielmehr die Wahrnehmung von abstrakten Eindrücken bestimmt, ob die Verbrauchenden etwas als authentisch wahrnehmen oder nicht (Morhart et al., 2015, S. 202). So können beispielsweise auch Replikationen, die nach objektivistischer Sicht nicht authentisch sein können, als authentisch wahrgenommen werden (z.B. S. Brown et al., 2003).

Wahrgenommene Markenauthentizität kann auch aus einer existentialistischen Perspektive betrachtet werden. Aus dieser Perspektive bedeutet Authentizität, sich selbst treu zu sein (Steiner & Reisinger, 2006, S. 300). Ein Objekt wird dann als authentisch angesehen, wenn es als identitätsbezogene Quelle dient. Für den Markenkontext bedeutet dies, dass eine Marke eine Ressource zur Offenbarung des eigenen Selbst sein sollte. Beim Konsum der Marke sollte das Gefühl entstehen, seinem Selbst treu zu sein (Morhart et al., 2015, S. 202).

Aus heutiger Sicht wird die Authentizität nicht strikt nach den drei genannten Perspektiven getrennt, sondern ergibt sich aus einer Verbindung dieser. Markenauthentizität ist demnach das Ausmaß „which consumers perceive a brand to be faithful and true toward itself and its consumers, and to support consumers being true to themselves“(Morhart et al., 2015, S. 202).

Die Faktoren, die bestimmen, wann oder wodurch eine Marke als authentisch wahrgenommen wird, unterscheiden sich in der Literatur. Je nach Autorenschaft werden unterschiedliche Dimensionen definiert, die zu wahrgenommener Markenauthentizität führen (z.B. Bruhn et al., 2012; Fritz et al., 2017; Morhart et al., 2015; Napoli et al., 2014). Ein Überblick verschiedener Dimensionen ist in Tabelle 1 zu sehen.

Tabelle 1: Übersicht Authentizitätsdimensionen (eigene Darstellung)

Bruhn et al. (2012)	Napoli et al. (2014)	Morhart et al. (2015)	Fritz et al. (2017)
- Kontinuität - Reliabilität - Natürlichkeit - Originalität	- Tradition/Erbe („Heritage“) - Ehrlichkeit - Qualitätsverpflichtung	- Kontinuität - Glaubwürdigkeit - Integrität - Symbolismus	- Vergangenheit einer Marke - Tugend einer Marke - Kulturelle Passung

Nachfolgend wird nur auf die vier Dimensionen Kontinuität, Integrität, Glaubwürdigkeit und Symbolismus eingegangen, die von Morhart et al. (2015, S. 202) ermittelt wurden.

Unter *Kontinuität* versteht man die Zeitlosigkeit einer Marke, die sich z.B. aus einer gewissen Historie entwickelt hat. Eine Marke ist dann *Glaubwürdig*, wenn sie das erfüllt, was sie verspricht. *Integrität* ist ein weiterer Aspekt, der zu wahrgenommener Authentizität beiträgt. Eine Marke ist dann integer, wenn sie moralische und ethische Werte hat und diese auch vertritt. Die letzte Dimension ist *Symbolismus*. Hierunter wird verstanden, dass Marken Werte widerspiegeln, die für das Selbstkonstrukt eines Individuums relevant sind (Morhart et al., 2015, S. 202f.).

2.3. Einstellung gegenüber einer Marke

Die Untersuchung der Einstellung gegenüber einer Marke ist aus mehreren Gründen interessant für die Forschung. Zum einen dient sie der Vorhersage des Konsumentenverhaltens (z.B. Faircloth et al., 2001; Mitchell und Olson, 1981). Andererseits ist es interessant, die theoretischen Ansätze zu betrachten, die die Entstehung der Einstellung (gegenüber einer Marke) erklären, um weitere Schlussfolgerungen zu ziehen (z.B. Ajzen und Fishbein, 1977; Petty und Cacioppo, 1986).

Allgemein wird die Einstellung gegenüber der Marke von Mitchell und Olson (1981, S. 381) als die interne Evaluation einer Marke seitens des Individuums gesehen. Spears und Singh (2004, S. 55) erweitern diese Sicht und definieren die Einstellung gegenüber der Marke als eine relativ beständige, eindimensionale zusammenfassende Bewertung der Marke, wodurch Verhalten angeregt wird.

Einer der am meisten untersuchten Einflussfaktoren auf die Einstellung im Markenkontext ist die Einstellung zur Werbung (Spears & Singh, 2004, S. 53). Demnach führt eine positive Einstellung gegenüber der Werbung auch zu einer positiveren Einstellung gegenüber der Marke (S. P. Brown & Stayman, 1992, S. 42f.). Dies findet konzeptionelle Fundierungen im Dual-Mediation Modell, welches genauer den Zusammenhang der Einstellung zur Werbung und zur Marke, Markenkenntnisse, und Kaufabsicht beschreibt (MacKenzie et al., 1986). Ein weiteres Modell, das sich mit der Einstellung gegenüber einer Marke beschäftigt, ist das Elaboration Likelihood Modell, welches erklärt wie sich persuasive Botschaften auf die Einstellung auswirken und sich zusätzlich mit der Stabilität der entstanden Einstellung beschäftigt (Petty & Cacioppo, 1981).

Um praktische Implikationen ableiten zu können, ist es besonders wichtig die Auswirkungen der Markeneinstellung zu kennen. Baldinger und Rubinson (1996, S. 33) zeigten in ihrer Forschung schon früh, dass sich durch die Einstellung zu einer Marke auch Verhalten voraussagen lässt. So fanden sie heraus, dass Markenloyalität maßgeblich von der Markeneinstellung abhängt. Dadurch lässt sich weiterhin erklären, dass bei zwei Drittel der von ihnen untersuchten Marken, der Marktanteil stieg, wenn die Einstellung zu Marke positiver wurde (Baldinger & Rubinson, 1996, S. 31). Die Einstellung zur Marke kann zudem als eine Art Markenassoziation betrachtet werden, wodurch sie ebenfalls als Teil des Markenimages angesehen werden kann. Das Markenimage beeinflusst wiederum den Markenwert. So besteht auch eine indirekte Beeinflussung des Markenwertes durch die Markeneinstellung (Faircloth et al., 2001, S. 70). Eine weitere, besonders für die Praxis relevante Auswirkung einer positiven Einstellung ist die Kaufabsicht. Personen, die eine bessere (vs. schlechtere) Einstellung zu einer Marke haben, haben eher (vs. weniger) die Absicht, Produkte der entsprechenden Marke zu kaufen (Spears & Singh, 2004, S. 61).

Jedes der zuvor genannten Konzepte könnte noch weiter vertieft und im Detail definiert werden. Hier dienen sie jedoch nur dazu, die Relevanz der Einstellung gegenüber einer Marke zu verdeutlichen. Zusammenfassend lässt sich demnach festhalten, dass die Einstellung gegenüber einer Marke maßgeblich zum Erfolg der Marke beiträgt und daher ein wichtiger Faktor ist, um ihn in Modelle miteinzubeziehen.

2.4. Involvement

Das Involvement-Konstrukt findet Anwendung in zahlreichen Forschungen, überwiegend um die Beziehung von Konsumierenden zu Werbung, zu Produkten oder auch zur Kaufentscheidung zu erklären (Zaichkowsky, 1985, S. 341). Grundsätzlich bezieht sich Involvement auf wahrgenommene Relevanz, beruhend auf inhärenten Bedürfnissen, Werten und Interessen (Zaichkowsky, 1985, S. 342). Eine Schwierigkeit bei der Nutzung des Involvement-Konstrukts besteht darin, dass es keine einheitliche Definition gibt. Es gibt verschiedene Arten, Antezedenzien und Konsequenzen von Involvement, die häufig nicht eindeutig voneinander abgegrenzt werden können und mit Definitionen verwechselt werden (Laurent und Kapferer, 1985, S. 42; Zaichkowsky, 1985, S. 341, Petty und Cacioppo, 1981).

Neben der Unterscheidung zwischen Involvement für Werbung, Produkte und Produktarten oder Kaufentscheidungen (Zaichkowsky, 1986), wird auch in langfristiges vs. situatives Involvement unterschieden. Während es sich beim situativen Involvement um ein kurzfristiges, meist anlassbezogenes Interesse handelt, beruht das langfristige Involve-

ment auf einem allgemeinen und dauerhaften Interesse und resultiert aus verankerten Werten (Laurent & Kapferer, 1985, S. 42).

Darüber hinaus wird in der Verbraucherpsychologie häufig zwischen Situationen mit hohem und niedrigem Involvement unterschieden (Petty & Cacioppo, 1981). Diese Perspektive ist Kern der vorliegenden Arbeit. Es besteht Einigkeit darüber, dass in Situationen mit hohem Involvement eine hohe persönliche Relevanz der Botschaft besteht. In Situationen mit niedrigem Involvement ist dagegen auch die Relevanz für das Individuum niedrig (Petty & Cacioppo, 1981). Hohes Involvement liegt vor, wenn eine intrinsische oder persönliche Bedeutung für die Situation besteht (Sherif et al., 1973, S. 311). Nach Krugman (1967, S. 355) werden die persönlichen Bezüge die ein Empfänger zwischen der Botschaft und dem eigenen Leben herstellen kann, als Involvement bezeichnet.

Es können drei Faktoren definiert werden, welche die Entstehung von Involvement maßgeblich beeinflussen. Als erstes wird es durch persönliche Faktoren beeinflusst. Dazu zählen alle Eigenschaften der Person, wie ihr Wertesystem, ihre Bedürfnisse, Interessen, Ziele und Erfahrungen. Als zweites wird Involvement durch Objekt- oder Stimulus- Faktoren beeinflusst. Hierzu gehören unter anderem die Quelle und der Inhalt der Botschaft. Als letztes sind die situativen Faktoren zu nennen. Eine Person, die erst kürzlich mit einem Thema in Berührung gekommen ist oder den Kauf eines bestimmten Produkts in Erwägung zieht, hat mit größerer Wahrscheinlichkeit ein höheres Involvement als eine Person, die keinen Kauf erwägt (Zaichkowsky, 1986, S. 5).

Haben Individuen ein hohes Involvement in einer bestimmten Situation, ist es wahrscheinlich, dass sie Informationen, vor allem persuasive Botschaften genauer verarbeiten (Petty & Cacioppo, 1981). Diese These findet theoretische Fundierung im Elaboration Likelihood Modell, dessen Grundannahme ist, dass Rezipienten persuasive Botschaften mit unterschiedlicher Intensität verarbeiten, was sich auf die Stabilität der Einstellungsänderung auswirkt (Petty & Cacioppo, 1981). Dabei meint Elaboration das Ausmaß, in dem eine Person über themenrelevante Argumente oder Informationen nachdenkt und diese kognitiv verarbeitet. Die Elaborationswahrscheinlichkeit ergibt sich aus der Motivation und Fähigkeit des Empfangenden Botschaften zu verarbeiten und wird durch das Involvement bestimmt (Petty & Cacioppo, 1986, S. 128).

Laut Elaboration Likelihood Modell gibt es zwei verschiedene Routen auf denen Informationen verarbeitet werden – die zentrale und die periphere Route. Man spricht von der zentralen Route, wenn die Elaborationswahrscheinlichkeit, also das Involvement hoch ist. Hierbei denkt der Konsument gründlich über die Argumente und Inhalte einer Nachricht nach und verbindet dies mit bereits vorhandenem Wissen. Die Einstellung ist demnach das Ergebnis dieser kognitiven Verarbeitung und beruht auf qualitativ überzeugenden Argumenten. Diese kognitive Verarbeitung der Informationen führt zu einer stabilen Einstellungsänderung (Petty & Cacioppo, 1986, S. 131). Ist die Wahrscheinlichkeit der Elaboration jedoch gering, also das Involvement niedrig, erfolgt die Verarbeitung neuer Informationen über die periphere Route. Die Einstellung wird eher emotional, oberflächlich und ohne aktive Auseinandersetzung mit den Argumenten gebildet. Sie ist nicht stabil und basiert auf peripheren Reizen wie der Attraktivität, der Kompetenz der Quelle oder der Anzahl der Argumente (Petty & Cacioppo, 1986, S. 125).

Das Elaboration Likelihood Modell erklärt demnach nicht nur die Auswirkung von Involvement auf die Einstellung zur Marke, sondern kann auch weiterführende Erkenntnisse darüber liefern, wann welche Werbung effektiv sein könnte und welche Argumente bei welchen Personen wirken können.

2.5. Skepsis

Häufig sind Konsumierende skeptisch gegenüber den Absichten von Marken. Besonders frequent wird dieser Aspekt in der Werbeforschung untersucht (z.B. Boush et al., 1994; Obermiller et al., 2005). Hier wird daran gezweifelt, ob die Werbebotschaft ehrlich ist und ob die Marke das einhält, was sie verspricht. Ein anderer relevanter Forschungszweig betrachtet die Skepsis gegenüber Unternehmenshandlungen (z.B. Forehand und Grier, 2003). Hier wird insbesondere untersucht aus welchem Grund CSR Aktivitäten angezweifelt werden, da ihre positive Wirkung dadurch abnimmt (z.B. Pirsch et al., 2007; Skarmeas und Leonidou, 2013). Die Untersuchung der Skepsis ist besonders in der Praxis relevant, da sie zur Erklärung beiträgt, wieso Bemühungen die Absatzzahlen (z.B. durch Werbung) oder das Image und die Loyalität (z.B. durch CSR Aktivitäten) zu erhöhen, nicht helfen (Boush et al., 1994; Obermiller et al., 2005; Skarmeas & Leonidou, 2013).

Nachfolgend wird näher auf Skepsis gegenüber Werbung bzw. allgemeiner Marketingbehauptungen und -aktivitäten eingegangen. Darunter wird eine Tendenz verstanden, diesen gegenüber misstrauischer und kritischer zu sein (Obermiller & Spangenberg, 1998, S. 160). Skepsis umfasst in diesem Zusammenhang nicht nur das Anzweifeln des Wahrheitsgehalts einer Werbeaussage, sondern auch die Kritik an den Motiven, dem Informationswert oder der Angemessenheit der Werbung für eine bestimmte Ziel- oder Produktgruppe (Obermiller & Spangenberg, 1998, S. 160). Es wird je nach Autorenschaft entweder als ein anhaltender Zustand (Boush et al., 1994; Obermiller & Spangenberg, 1998) oder als situativer Zustand, der vom Kontext abhängt, gesehen (Mohr et al., 1998; Skarmeas & Leonidou, 2013).

Ein verwandter Begriff, der oft mit Skepsis gleichgesetzt wird, ist Zynismus. Zynische Menschen denken, dass grundsätzlich aus egoistischen Motiven gehandelt wird und stellen das Wohlwollen anderer stark in Frage (Kanter & Wortzel, 1985, S. 6). Zynismus wird als allgemeine Charaktereigenschaft angesehen, während Skepsis gegenüber Werbung sich eher aus den persönlichen Markterfahrungen ergibt und demnach „angelernt“ ist (Obermiller & Spangenberg, 1998, S. 166). Ein zentraler Unterschied der beiden Begriffe besteht darin, dass Skepsis dadurch gekennzeichnet ist, dass sie entkräftet werden kann bzw. Menschen mit ausreichenden Informationen und Beweisen überzeugt werden können (Mohr

et al., 1998, S. 33). Aspekte wie die Eigenschaft der Quelle, persönliches Vorwissen und andere Variablen der Botschaft, die die Überzeugungskraft von Informationen beeinflussen können, wirken sich ebenfalls auf die Akzeptanz der Marketingbotschaft aus (Obermiller & Spangenberg, 1998, S. 160).

Personen, die skeptisch gegenüber Werbung sind, sind sich der persuasiven Natur dieser bewusst und entwickeln Wissen darüber. Dieses Wissen hilft zu erkennen, wann man beeinflusst wird. So ist es für Personen einfacher, die Motive, Strategien und Taktiken der VermarkterInnen zu erkennen (Campbell & Kirmani, 2000, S. 70). Dies hat zur Folge, dass sich insgesamt weniger auf die Werbung verlassen wird, ihr weniger Beachtung geschenkt wird und allgemein, dass die Botschaft weniger positiv wahrgenommen wird (Campbell & Kirmani, 2000, S. 81).

3. Hypothesenherleitung

3.1. Konzeptuelles Modell

Nach der theoretischen Fundierung der fünf Variablen, aus denen sich das konzeptuelle Modell (Abb. 1) zusammensetzt, werden nun die einzelnen Zusammenhänge grundlegend erklärt und anschließend die Hypothesen im Detail hergeleitet. Der untersuchte Haupteffekt ist der Einfluss von authentischem Markenaktivismus auf die Einstellung gegenüber der Marke. Somit stellt authentischer (vs. nicht authentischer) Markenaktivismus die unabhängige Variable und die Einstellung gegenüber der Marke die abhängige Variable dar. Es wird angenommen, dass die wahrgenommene Markenauthentizität diesen Effekt mediiert. Zusätzlich wird überprüft, ob die Moderatoren Involvement in das Thema Rassismus und Skepsis gegenüber Werbung einen Einfluss darauf haben, ob authentischer Markenaktivismus auch als solcher erkannt wird und zu einer größeren wahrgenommenen Markenauthentizität führt. Daraus ergibt sich das konzeptuelle Modell in Abb. 1.

3.2. Der Einfluss von authentischem Markenaktivismus auf die Einstellung gegenüber der Marke

Wird Markenaktivismus als authentisch angesehen, sollte dies zu einer positiven Einstellung gegenüber der Marke führen. Eine Argumentationsgrundlage für diesen Zusammenhang ist die Identifikation von Konsumierenden mit der Marke (Stokburger-Sauer et al., 2012, S. 409). Marken dienen als wichtige Ressource für die Identität der Konsumierenden und werden dafür genutzt, um das eigene Selbst auszudrücken, es zu bestätigen oder aufzuwerten (Aaker, 1997, S. 347). Wenn man Marken konsumiert, die zur eigenen Identität passen, bzw. mit denen man sich identifizieren kann, ist es möglich eine Version des eigenen Selbstkonzepts nach außen hin ausdrücken, wodurch sich dieses manifestiert (Reed et al., 2012, S. 310). Insbesondere Authentizität ist sehr wichtig für das Selbstkonzept eines Menschen und wird daher dringend angestrebt (Napoli et al., 2014, S. 1090).

Aus der Theorie zur Identifikation von Verbrauchenden mit Marken geht hervor, dass es verschiedene Aspekte gibt, die zu einer höheren Identifikation mit der Marke führen (Stokburger-Sauer et al., 2012, S. 409). Vorherige Forschung im Kontext von Markenaktivismus fand heraus, dass besonders die erhöhte Ähnlichkeit zu einer Marke zu einer größeren Identifikation führt und demnach ausschlaggebend für positive Marketingergebnisse wie z.B. der Kaufabsicht ist (Hydock et al., 2020; Mukherjee & Althuizen, 2020).

Authentischer Markenaktivismus spiegelt die pro-sozialen Werte des Unternehmens nach außen wider und ist im Sinne des Zwecks ausgerichtet, statt eigene Vorteile anzustreben (Vredenburg et al., 2020, S. 451). Da das nach außen vermittelte Bild mit den inneren Werten zusammenpasst, wirkt der Markenaktivismus aufrichtig und authentisch (Cinelli & LeBoeuf, 2020, S. 41). Dies sollte dazu führen, dass Individuen eine größere Ähnlichkeit in Bezug auf Authentizität zwischen ihrem Selbstkonzept und der Marke wahrnehmen. Folglich sollte eine bessere Einstellung gegenüber der Marke entstehen, ausgelöst durch die größere Identifikation mit der Marke.

Ein weiterer Aspekt, der zu einer größeren Identifikation mit der Marke führt ist Markenwärme (Stokburger-Sauer et al., 2012, S. 407). Authentischem Markenaktivismus können durch die vertretenen pro-sozialen Werte und dem sozialen Engagement solche warmen Attribute zugeschrieben werden. Auch nach dieser Argumentation kann es zu einer stärkeren Identifikation mit der Marke und der damit verbundenen Einstellungsverbesserung kommen.

Ebenso wie Konsumierende Marken mögen, die dem eigenen Identitätskonzept entsprechen, sind sie abgeneigt gegenüber Marken, die nicht zu ihrer Identität passen (White et al., 2012, S. 704). Es kann davon ausgegangen werden, dass Individuen aufgrund des Mangels an Authentizität eine Abneigung gegen nicht authentischen Markenaktivismus haben.

Die fehlende Passung der Botschaft, der Unternehmenspraktiken, der Werte und des Purpose, die zu nicht authentischem Markenaktivismus führt, kann durch verschiedene Kombinationen der Ausprägungen dieser Aspekte entstehen. Es handelt sich beispielsweise dann um nicht authentischen Markenaktivismus, wenn pro-soziale Werte und eine aktivistische Botschaft vorhanden sind, ohne dass ein ernsthafter Versuch unternommen wird, das soziale Problem zu lösen. Es wird aber auch als unauthentisch angesehen, wenn auf die aktivistische Botschaft scheinbar ernsthafte Versuche folgen, das Problem zu lösen, die Marke aber gegenteilige nichtsoziale Werte vertritt. So könnten hier je nach Kombination verschiedene Effekte beobachtet werden. Dennoch wird erwartet, dass nicht authentischer Markenaktivismus aufgrund der mangelnden Konsistenz des Handelns und unklaren Absichten als weniger authentisch angesehen wird, wodurch auch eine geringere Identifikation mit der Marke bestehen sollte. Insgesamt sollte sich nicht authentischer Markenaktivismus demnach im Vergleich zu authentischem Markenaktivismus negativer auf die Einstellung gegenüber der Marke auswirken.

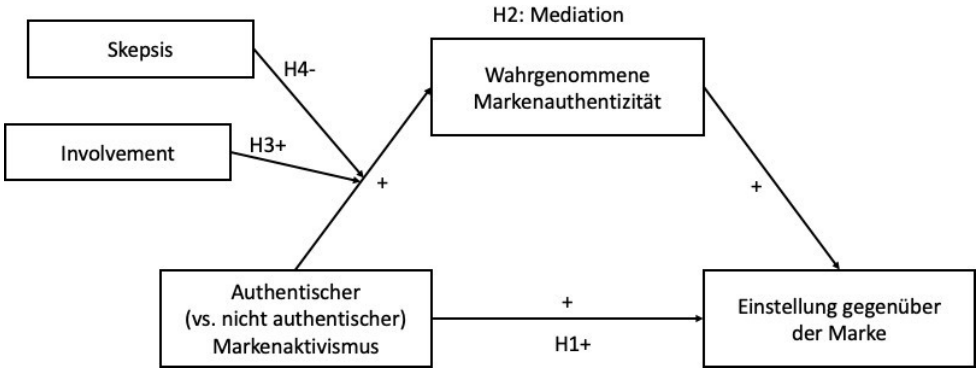


Abbildung 1: Konzeptuelles Modell (eigene Darstellung)

Deshalb lautet die Hypothese:

H1: Authentischer (vs. nicht authentischer) Markenaktivismus führt zu einer positiveren (negativeren) Einstellung gegenüber der Marke.

3.3. Der mediierende Effekt von wahrgenommener Markenauthentizität

Das Vorhandensein der verschiedenen Perspektiven von Authentizität (Morhart et al., 2015, S. 201) zeigt, dass Authentizität ein komplexes und nicht eindeutiges Konstrukt ist, das genauer betrachtet werden muss. Da Authentizität nicht rein objektiv ist (Grayson & Martinec, 2004, S. 298), ist es nicht zwangsweise so, dass authentischer Markenaktivismus auch als solcher empfunden wird. Vor allem die konstruktivistische Sicht lässt erwarten, dass Markenaktivismus von Individuum zu Individuum unterschiedlich bewertet wird, was zum Beispiel an unterschiedlichen Ansichten oder Erwartungen liegen kann (Wang, 1999, S. 351). Auf der anderen Seite wäre es möglich, dass Konsumierende durch mangelndes Interesse oder Aufmerksamkeit nicht erkennen können, was authentisch ist und was nicht. An dem in der Einleitung vorgestellten Beispiel der Marke Nike, die eine Werbekampagne mit dem ehemaligen Footballspieler Colin Kaepernick zeigten, kann man sehen, dass es möglicherweise zu unterschiedlichen Bewertungen der Authentizität kommen könnte. Die vermittelte Botschaft würde wohl von den meisten als stimmig mit den Werten von Nike angesehen werden, wodurch das Engagement zunächst authentisch scheint. Fraglich ist, wie die zusätzliche Information, bewertet wird, dass Nike weiterhin die Mannschaft sponsorte, die die Proteste von Colin Kaepernick verboten hat (Carp, 2018). Es muss persönlich bewertet werden, ob die beiden Handlung im Gegensatz zueinander stehen und der Markenaktivismus demnach nicht mehr als authentisch gesehen wird oder ob man die Meinung vertritt, dass beide Handlungen parallel ethisch vertretbar sind. Dies setzt jedoch grundsätzlich voraus, dass die Information bei den Konsumierenden vorhanden ist.

Ob die zuvor erläuterten Definitionen von authentischem und nicht authentischem Markenaktivismus auch in der Realität abgebildet werden können und Marken dadurch authentischer bzw. nicht authentischer wirken, muss demnach zusätzlich überprüft werden. Dafür wird die wahrgenommene Markenauthentizität als Mediator ins Modell aufgenommen. Nur so kann sichergestellt werden, dass die Effekte auch auf die Authentizität und nicht auf andere Aspekte zurückzuführen sind.

Grundsätzlich ist zu erwarten, dass authentischer Markenaktivismus, durch die Stimmigkeit der Werte, Unternehmenspraktiken, Botschaft und dem Purpose dazu führt, dass auch die Marke als authentischer wahrgenommen wird. Im Umkehrschluss sollte nicht authentischer Markenaktivismus dazu beitragen, dass die Marke weniger authentisch wirkt. Es wird erwartet, dass die Authentizität ein Haupttreiber für die Verbesserung der Einstellung ist.

Zusammengefasst wird demnach erwartet, dass durch authentischen Markenaktivismus auch die Marke authentischer wahrgenommen wird, wodurch es zu einer besseren Einstellung kommt.

Dementsprechend lautet die Hypothese:

H2: Der Effekt von authentischem Markenaktivismus auf die Einstellung zur Marke wird mediiert durch die wahrgenommene Markenauthentizität.

3.4. Die moderierende Wirkung von Involvement und Skepsis

Es wird davon ausgegangen, dass wahrgenommene Markenauthentizität den Effekt von authentischem Markenaktivismus auf die Einstellung gegenüber der Marke mediiert, und die gesteigerte wahrgenommene Markenauthentizität der Haupttreiber für eine Einstellungsverbesserung ist. Deshalb ist es vor allem interessant herauszufinden, was die Beziehung zwischen der unabhängigen Variable authentischer Markenaktivismus und dem Mediator wahrgenommene Markenauthentizität beeinflusst. Was könnten Gründe dafür sein, dass authentischer Markenaktivismus nicht zu einer größeren wahrgenommenen Markenauthentizität führt? Durch welche Faktoren könnte sich der positiv erwartete Effekt erhöhen? Um Antworten auf diese Fragen zu finden, wird überprüft, ob das Involvement in Bezug auf das gesellschaftlich diskutierte Thema oder Skepsis gegenüber Werbung bzw. Marketingaktivitäten, solche Einflussfaktoren sind.

Skepsis

Skepsis gegenüber Werbung, wurde im Kontext wahrgenommener Markenauthentizität bereits von Morhart et al. (2015, S. 212) empirisch untersucht. Sie fanden heraus, dass Verbrauchende besonders dann die Authentizität anzweifeln, wenn sich das Marketing auf schwierig nachweisbare Aspekte wie z.B. Werte ausrichtet. Auch im Kontext von Markenaktivismus müssten die Werte der Marke näher betrachtet werden, um das dem Engagement zugrundeliegende Motiv besser bewerten zu können.

Ein großes Problem besteht darin, dass einige Marken Woke Washing betreiben, was sich negativ auf das allgemeine Bild von Markenaktivismus auswirkt (Vredenburg et al., 2020, S. 449). So sind Menschen oft skeptisch, über die Aufrichtigkeit der Marke (Edelman, 2019, S. 14; Skarmeas und Leonidou, 2013, S. 1831f.). Die hierbei präsente Frage ist, ob Marken wirklich für den aktivistischen Standpunkt einstehen oder ob dem Engagement lediglich opportunistische Gedanken zugrunde liegen.

In welchem Maß Markenaktivismus als Werbung bzw. Marketingmaßnahme angesehen werden kann, hängt stark von der Form ab. Im wünschenswerten Fall steht primär das Engagement für ein sozio-politisches Thema im Vordergrund und nicht das Bewerben eines Produktes. In einigen Fällen enthält der Markenaktivismus aber auch eine verkaufsfördernde Komponente. Es ist zum Beispiel häufig zu beobachten, dass das Einstehen für die LGBTQ+ Community mit dem Verkauf einer „Pride“ Kollektion kombiniert wird. In diesem Fall ist die Werbemaßnahme sehr präsent. Formunabhängig kann Markenaktivismus jedoch, eine duale Natur unterstellt werden – das Engagement kann sowohl ursachennützlich als auch ursachenausbeutend sein (Skarmeas & Leonidou, 2013, S. 1832). Als ursachennützlich wirkt Markenaktivismus dann, wenn der Öffentlichkeit dienende Motive im Vordergrund stehen (Forehand & Grier, 2003, S. 350). Sofern das Ziel des Markenaktivismus ist, eine Lösung des gesellschaftlich relevanten Problems zu finden, kann es als ursachennützlich gesehen werden.

Markenaktivismus kann zudem als ursachenausbeutend gesehen werden. In diesem Fall würden dem Handeln vor allem firmenbezogene Motive unterstellt werden (Forehand & Grier, 2003, S. 350), was allgemein in Bezug auf Markenaktivismus, als das Ausnutzen des Problems für den eigenen Profit gesehen werden kann. Bei nicht authentischem Markenaktivismus sind diese Aspekte besonders ausgeprägt. Aber auch Marken die ernsthaft die Förderung der Problemlösung vorantreiben, können firmenbezogene Motive unterstellt werden. Es könnte unterstellt werden, dass der Markenaktivismus implementiert wurde, damit die Marke besonders engagiert wirkt und ihr Attribute wie sozial, warm und meinungsstark zugeschrieben werden.

Ist eine Person nun skeptischer und hat somit bereits Wissen über Marketingmechanismen erlangt, ist sie sich der dualen Natur von Markenaktivismus bewusst. So werden wahrscheinlicher auch die ursachenausbeutenden Aspekte genauer betrachtet. Die vorhandene Skepsis führt dazu, dass die wahren Absichten der Marke insgesamt kritischer hinterfragt werden, weniger Aufrichtigkeit dahinter vermutet wird (Campbell & Kirmani, 2000, S. 81) und vermutlich auch die Authentizität als geringer bewertet wird.

Ist eine Person hingegen weniger skeptisch, wird sie den Aussagen der Marke eher glauben und weniger schlechte Hintergedanken bezüglich der Motive der Marke haben. So wird der authentische Markenaktivismus weniger stark angezweifelt und die wahrgenommen Markenauthentizität besser bewertet.

So ergibt sich folgende Hypothese:

> *H3: Wenn Skepsis gegenüber Werbung hoch (niedrig) ist, dann ist der positive Effekt von authentischem Markenaktivismus auf die wahrgenommene Markenauthentizität schwächer (stärker).*

Involvement

Die Betrachtung von Involvement als Moderator ist vor allem dadurch interessant, da Markenaktivismus teilweise zu sehr extremen Reaktionen führt, wie z.B. zum Boykott der Marke (Norman, 2012). Demnach scheint es unterschiedliche persönliche Relevanzen bezogen auf die vom Markenaktivismus behandelten Themen zu geben.

Ist das Involvement zum Thema des Markenaktivismus (hier Rassismus) hoch, ist auch die Situation von höherer Bedeutung. Dadurch werden die Informationen auf der zentralen Route verarbeitet. Das heißt, die Informationen werden intensiver und anhand qualitativer Argumente verarbeitet (Petty & Cacioppo, 1986, S. 131). So sollten die wahren Absichten des Unternehmens bezogen auf den Markenaktivismus eher erkannt werden. Da auf der zentralen Route bereits vorhandenes Wissen abgerufen und verknüpft wird (Petty & Cacioppo, 1986, S. 131), sollte es Teilnehmende mit hohem Involvement so leichter fallen die einzelnen Bausteine die zu authentischem Markenaktivismus führen zu erkennen. Sie sollten sich demnach bei der Betrachtung des Markenaktivismus auch über vorhandene Informationen zu den Werten der Marke bewusst werden und bewerten können, ob sie einen Purpose hat. Die Informationen sollten verknüpft werden und es sollte deutlich werden, wann sich eine Marke auf eine authentische Weise bemüht und wann nicht. Zudem sollte es für hoch involvierten Personen durch das wahrscheinlich vorhandene Vorwissen zum Thema (Petty & Cacioppo, 1986, S. 131) leichter sein zu beurteilen, ob der Markenaktivismus auch wirklich hilfreich ist oder ob es nur den Anschein erwecken soll.

Im Vergleich zu hohem Involvement und der Verarbeitung auf der zentralen Route fehlt bei geringem Involvement die Motivation, sich näher mit dem Thema auseinanderzusetzen, um sich anschließend eine fundierte Meinung über die Authentizität des Markenaktivismus zu bilden. Die Verarbeitung der Informationen findet auf der peripheren Route statt. Hier werden die Information lediglich oberflächlich und anhand von nicht inhaltlichen Merkmalen bewertet (Petty & Cacioppo, 1986, S. 125). Aspekte wie die Werte oder Purpose der

Marke werden somit nicht berücksichtig. Dadurch fehlt eine wichtige Komponente um den Markenaktivismus als authentisch zu bewerten, was sich dementsprechend auch auf die wahrgenommene Markenauthentizität auswirkt.

Deshalb lautet die Hypothese:

> *H4: Wenn Involvement hoch (niedrig), dann ist der positive Effekt von authentischem Markenaktivismus auf die die wahrgenommene Markenauthentizität stärker (schwächer).*

4. Empirische Studie

4.1. Methodisches Vorgehen

4.1.1. Durchführung

Um die Hypothesen zu testen, wurde ein 2x2 faktorielles Experimentaldesign gewählt. Das Experiment wurde in Form einer Online-Umfrage auf der Plattform Tivian durchgeführt. Ziel des Experiments ist es, herauszufinden, welche Rolle Authentizität bei der Betrachtung von Markenaktivismus hat. Dafür wurde als erstes der Einfluss von authentischem Markenaktivismus auf die Einstellung zur Marke getestet und danach, ob dieser durch die wahrgenommene Markenauthentizität mediiert wird. Als drittes wurde zusätzlich geprüft, ob die Beziehung zwischen authentischem Markenaktivismus und wahrgenommener Authentizität von den Moderatoren Skepsis gegenüber Werbung und Involvement in das vom Markenaktivismus behandelte Thema beeinflusst wird. Dafür wurde die Variable authentischer Markenaktivismus systematisch in zwei Schritten manipuliert. Die Einstellung gegenüber der Marke, die wahrgenommene Markenauthentizität, Involvement und Skepsis wurden anhand von Skalen aus der Literatur abgefragt.

Vor Beginn der Befragung wurde ein Pretest mit 12 Teilnehmenden durchgeführt. Während acht Personen den Pretest ausfüllten und die Möglichkeit hatten Kommentare zu Verständnisproblemen zu hinterlassen, wurde mit den anderen vier Personen ein Think-Aloud-Pretest durchgeführt. Ziel war es, Probleme und Missverständnisse aufzudecken, um so die Qualität des Fragebogens zu gewährleisten. Besonderes Augenmerk lag dabei auf die Verständlichkeit und Wirksamkeit der Manipulationen. Nach der Einarbeitung des Feedbacks startete die Befragung. Sie fand einmalig statt und erstreckte sich über einen Zeitraum vom 14.07.2022 bis zum 30.07.2022. Die Dauer der Befragung war auf 5-7 Minuten ausgelegt. Alle Befragten hatten die Möglichkeit zwischen den Sprachen Englisch und Deutsch zu wählen. Bei der Stichprobe handelt es sich um ein Gelegenheitsstichprobe, da die Umfrage hauptsächlich über verschiedene Social-Media-Kanäle und im persönlichen Freundeskreis verbreitet wurde. Bevor die Operationalisierung der einzelnen Variablen in Kapitel 4.1.2 näher erläutert wird, wird nachfolgend der Aufbau des Fragebogens kurz beschrieben. Eine Übersicht des Aufbaus ist in Abb. 2 zu sehen.

Zuerst wurden die Moderatoren, das Involvement zum Thema Rassismus und die Skepsis gegenüber Werbung abgefragt. In dieser Studie soll der Einfluss von Authentizität auf die Beziehung von Markenaktivismus und der Einstellung gegenüber der Marke untersucht werden. Da aus anderen Studien bekannt ist, dass die Übereinstimmung mit dem Standpunkt der Marke ein Hauptbeeinflussungsfaktor dafür ist, ob sich Markenaktivismus positiv oder negativ auswirkt, wurde ein möglichst unkontroverses Thema für den Markenaktivismus gewählt. Ziel war es den Faktor „Übereinstimmung mit dem Standpunkt des Marke“ konstant zu halten. Dies wurde mit der Kontrollfrage, inwiefern man eine Bewegung gegen Rassismus unterstützt, sichergestellt. Anschließend bekamen die Teilnehmenden einen kurzen Einleitungstext über die fiktive Modemarke Ja&Do Fashion zu sehen, damit die Teilnehmenden sich die Marke besser vorstellen konnten. Im nächsten Schritt wurde den Teilnehmenden je nach Experimentalgruppe ein positives oder ein negatives Arbeitgeberranking gezeigt. Anschließend bekamen sie in der zweiten Manipulation einen Zeitungsartikel, in dem zum einen die Förderung des Problems und zum anderen opportunistisches Handeln im Fokus standen. Dies ergibt insgesamt vier Versuchsgruppen. Die Aufteilung in die verschiedenen Gruppen fand mittels Randomisierung statt. Die genaue Manipulation wird im nächsten Teilkapitel erklärt. Nach beiden Manipulation wurde jeweils die Einstellung ermittelt. Da die Manipulation der Werte der Marke im ersten Schritt einen erheblichen Einfluss auf die Bewertung der Einstellung hat, wird die Einstellungsänderung betrachtet. Das heißt, es wird untersucht, ob sich die positive bzw. negative Einstellung, welche auf Grundlage der Werte und der Mission der Marke gebildet wurde, in Abhängigkeit des Folgeszenarios ändert. Um zu überprüfen, ob die Manipulation gewirkt hat wurde jeweils ein Manipulationscheck durchgeführt. Nach den Manipulationen und der Einstellungsabfrage wurden der Mediator, die wahrgenommene Markenauthentizität und demografische Daten wie Alter, Geschlecht und die Tätigkeit abgefragt.

Um die Ergebnisse auszuwerten, werden zunächst die Manipulationschecks ausgewertet, die Reliabilität der Skalen überprüft und anschließend die Hypothesen mit geeigneten Analyseverfahren getestet.

4.1.2. Operationalisierung

Authentischer vs. nicht authentischer Markenaktivismus

Da authentischer Markenaktivismus nach der Definition von Vredenburg et al. (2020) getestet werden soll, erfolgt die Manipulation hierfür in zwei Schritten. In der ersten Manipulation werden die Werte und die Mission Marke manipuliert. Hieraus soll ein Szenario entstehen, dass eine Marke mit sozialen Werten zeigt, die eine pro-soziale Mission verfolgt. Das gegensätzliche Szenario, soll zeigen dass weder soziale Werte vorhanden sind noch Anzeichen dafür gibt, dass Ziele verfolgt werden, die über das unternehmerische Interesse hinausgehen.

Um die Werte der Marke glaubhaft und realistisch darzustellen, wird die Form einer Arbeitgeberbewertung gewählt. Die Teilnehmenden sehen die Bewertungen verschiedener

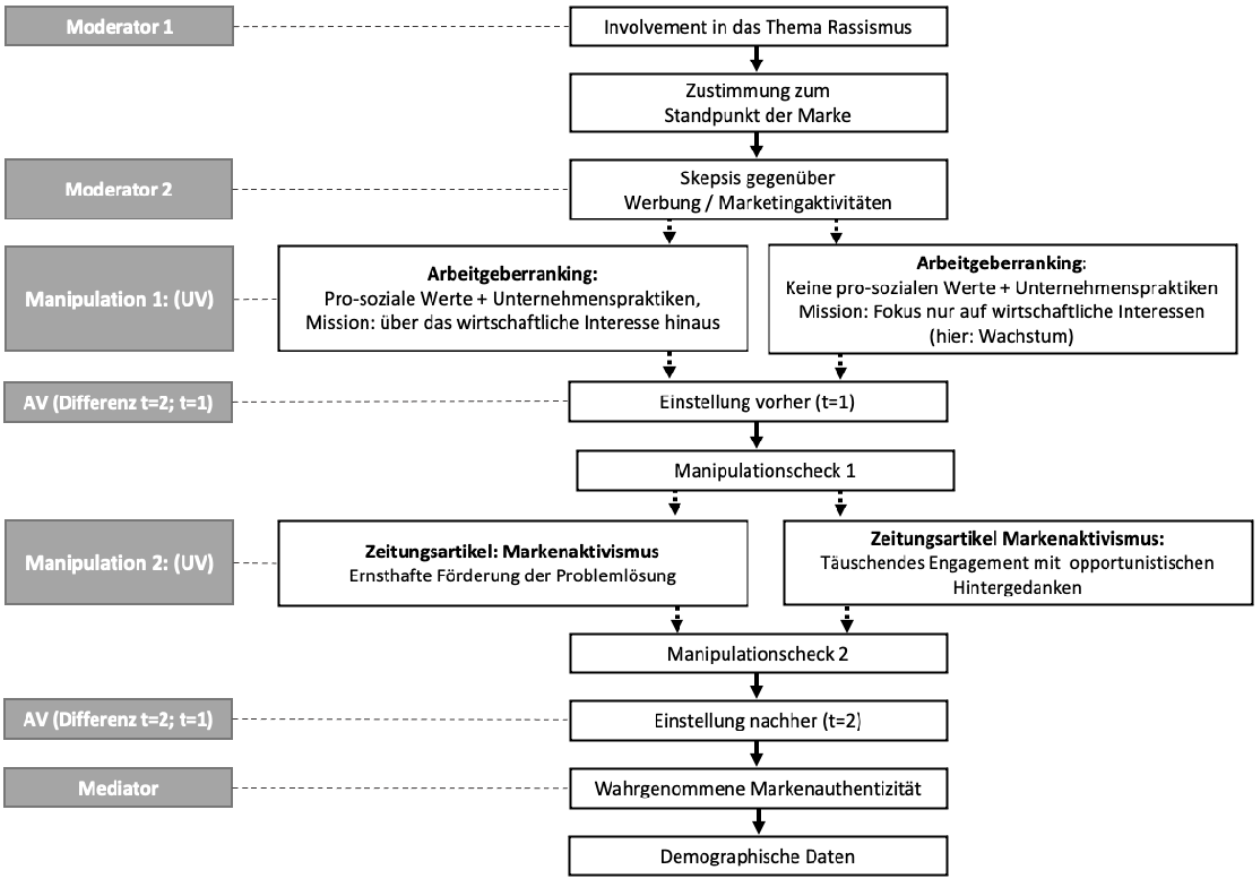


Abbildung 2: Aufbau Fragebogen (eigene Darstellung)

Aspekte, wie Work-Life-Balance, Umwelt- und Sozialverhalten, Diversität, Gleichberechtigung und Gehalt. Bei der Auswahl der Aspekte wurde das bekannte Bewertungs-Portal Kununu zur Orientierung herangezogen (Kununu, 2022). Aus diesen Aspekten lassen sich die vertretenen Werte ableiten. Wenn beispielsweise eine hohe Bewertung für das Umweltverhalten zu sehen ist, lässt sich daraus schließen, dass die Unternehmenspraktiken, die dieses Verhalten fördern, bereits in gewissem Maße vorhanden sind. Das Verwenden einer Arbeitgeberbewertung ist einerseits leicht zu verstehen und findet Anwendung in der Realität. Zusätzlich ist diese Methode glaubhaft, da die Information nicht vom Unternehmen selbst kommt (Du et al., 2010, S. 13). Dabei unterscheiden sich die Szenarien vor allem in der Höhe der Bewertung der sozialen Aspekte wie Diversität, Gleichberechtigung, Umwelt- und Sozialverhalten und Arbeitsbedingungen. Die Bewertung wurde um eine Information über die Vision des Unternehmens ergänzt, aus der sich ableiten lässt, ob die Marke von einem, Purpose geleitet wird oder nicht. Während das Szenario mit der sozialen Ausprägung die Mission sehr auf soziale und nachhaltige Aspekte ausrichtet, fokussiert sich das gegensätzliche Szenario lediglich auf Wachstum, ohne soziale Aspekte zu benennen.

Es ist wichtig, dass bei der ersten Bewertung der Marke nicht zu Beginn bereits die Endpunkte der Skala erreicht werden, damit es möglich ist eine Verbesserung bzw. Verschlechterung zu ermitteln. Aus diesem Grund wurden bei beiden Szenarien ebenfalls negative bzw. positive Aspekte eingebaut.

Der Manipulationscheck fragt ab, inwieweit dem Unternehmen soziale Werte und Motive zugeschrieben werden.

Im zweiten Schritt wird der Markenaktivismus manipuliert. Hier soll ein Szenario das Gefühl erwecken, dass das Problem ernsthaft unterstützt wird. Aus dem anderen Szenario sollen die opportunistischen Hintergedanken der Marke hervorgehen. Die aktivistische Botschaft ist bei beiden Manipulationen gleich und nur die damit einhergehenden Handlungen werden verändert.

Hierfür wird ein Zeitungsartikel als Informationsquelle genutzt. Es wird von externer Stelle berichtet, dass die Marke sich für eine Anti-Rassismus Bewegung einsetzt. Der Artikel besteht aus drei Komponenten. Zunächst wird in beiden Varianten erwähnt, dass die Proteste unterstützt werden und das Logo der Bewegung zu ihrer Online Präsenz hinzugefügt wurde. Als zweites ist ein Statement des CEOs zu sehen. Beide Bekunden im Namen der Marke ihre Solidarität und dass ihnen das Thema am Herzen liegt, da sie sich „jeher für Werte wie Diversität und Gleichberechtigung“ eigesetzt haben. In Kombination mit den nicht sozialen Werten würde dies eine Lüge darstellen. Als letzter Punkt wur-

de in das aktuelle organisatorische Handeln eingeschritten. In dem Szenario welches ehrliches Interesse wiederspiegeln sollte, wird eine Spende getätigt, sowie die aktuelle Werbekampagne gestoppt und die Werbezeit, stattdessen den AktivistInnen überlassen. Bei dem konträren Szenario, wird die aktuelle Kampagne ebenfalls gestoppt und stattdessen eine Stoppt-Rassismus-Kollektion in Verbindung mit einer Rabattaktion herausgebracht. Es wird nicht deutlich, dass über die Solidaritätsbekundung heraus etwas getan wird, um die Proteste zu unterstützen oder nachhaltig das Problem zu verbessern. Stattdessen soll der Anschein erweckt werden, dass die Marke sich nur engagiert, um selbst Vorteile zu erlangen. Um die Manipulation zu überprüfen wurde ein objektiver Manipulationscheck ergänzt, der gleichzeitig auch als Aufmerksamkeitscheck dient. Die Teilnehmenden werden gefragt, was sie im zuvor gezeigtem Artikel gelesen haben. So kann festgestellt werden, ob die Teilnehmenden vom richtigen Ausgangsszenario ausgehen.

Durch die Kombination der beiden Manipulationen entstehen vier Gruppen, von denen lediglich die Gruppe prosoziale Werte, Unternehmenspraktiken und Mission, und aufrichtige Förderung der Problemlösung zu authentischem Markenaktivismus führen sollte. Die übrigen drei Gruppen gelten laut Definition als nicht authentisch (s. Abb. 3).

Einstellung gegenüber der Marke

Die Einstellung gegenüber der Marke wird mit einer 3-Item Skala gemessen, die in vielen Forschungen Anwendung findet (z.B. Ewing et al., 2012; Nan und Heo, 2007). Die Befragten sollen auf einem semantischen Differential angeben, ob sie die Marke positiv oder negativ finden, sie mögen oder nicht mögen, oder ob sie ihnen gefällt bzw. nicht gefällt. Die Fragen werden jeweils nach der Manipulation gestellt. Betrachtet man im Nachhinein die Differenz der beiden Einstellungsmessungen, erhält man die Einstellungsänderung. So können insbesondere auch die Einstellungsänderungen der Szenarien untersucht werden, bei denen die Valenz der beiden Manipulationen in unterschiedliche Richtungen geht.

Wahrgenommene Markenauthentizität

Um die wahrgenommen Markenauthentizität zu messen wurde die validierte Skala von Morhart et al. (2015) genutzt. Diese besteht in ihrer ursprünglichen Form aus 15 Items, die in vier Subskalen unterteilt sind – Integrität, Symbolismus, Glaubwürdigkeit und Kontinuität. Für diese Arbeit wurden nur die ersten drei Subskalen verwendet. Die Dimension Kontinuität wurde ausgelassen, da sich dieser Aspekt der wahrgenommen Markenauthentizität auf die Historie des Unternehmens und die Beständigkeit der Marke über einen längeren Zeitraum bezieht. Da es sich in diesem Experiment um eine fiktive Marke handelt gibt es keine Informationen die diese Punkte behandeln. Deshalb ist es für die Teilnehmenden nicht möglich die Fragen zu dieser Dimension zu beantworten. Die final verwendete Skala für die wahrgenommene Markenauthentizität besteht demnach aus elf Items zu den Authentizitätsdimensionen Integrität, Symbolismus und Glaubwürdigkeit. Die Items wurden auf einer Siebener-Likert-Skala von „stimme gar nicht zu“ bis „stimme voll und ganz zu“ gemessen.

Analog zur Arbeit von Cinelli und LeBoeuf (2020) wird zur Kontrolle die wahrgenommene Markenauthentizität zusätzlich mit einem einzigen Item abgefragt.

Involvement

Um das Involvement-Konstrukt zu messen wurde die Skala von von Wangenheim und Bayón (2007) genutzt, welche wiederum auf der Forschung von Zaichkowsky (1985) basiert. Um die Skala im Kontext von Rassismus zu verwenden, wurden die Items dahingehenden angepasst. Es wurde dabei darauf geachtet, dass dennoch der Sinn der Ursprungsitems beibehalten wurde. Außerdem wurde die Skala gekürzt. So wird Involvement mit vier Items, die das Interesse und das Wissen über das Thema Rassismus, ob man sich regelmäßig über das Thema informiert und mit seinen Freunden darüber spricht, abgefragt. Die Items wurden ebenfalls auf einer Siebener-Likert-Skala von „stimme gar nicht zu“ bis „stimme voll und ganz zu“ gemessen.

Skepsis gegenüber Werbung

Um den Moderator Skepsis gegenüber Werbung abzufragen wurde die Skala von Gaski und Etzel (1986) verwendet. Diese Skala wurde insbesondere im Zusammenhang mit wahrgenommener Markenauthentizität bereits von Morhart et al. (2015) verwendet. Es wird beispielweise gefragt, ob man Werbung mag, sie nervig findet, ob man findet, dass sie den Konsumierenden täuscht und ob man ohne sie besser dran wäre. Auch hier wurden die Items auf einer Siebener-Likert-Skala von „stimme gar nicht zu“ bis „stimme voll und ganz zu“ gemessen.

4.2. Ergebnisse

Zur Vorbereitung der Analyse wurde der Datensatz bereinigt und auf die Güte überprüft. Hierfür wurden zunächst die Manipulationscheks ausgewertet. Alle Teilnehmenden, die diese nicht bestanden haben, wurden aus der Analyse ausgeschlossen, da davon auszugehen ist, dass ihre Ergebnisse nicht auf den Manipulationen beruhen. Für den ersten Manipulationschek heißt das, dass nur die Teilnehmenden, die die Manipulation für soziale bzw. nicht soziale Werte und Motive, auch dementsprechend bewertet haben, weiterhin betrachtet wurden. Da dies auf einem semantischem Differential mit sieben Schritten (1 = keine sozialen Werte und Motive, 7 = sehr ausgeprägte soziale Werte und Motive) gemessen wurde, galten die Werte 1 und 2 als untere Grenze, die Werte 3 bis 5 als mittlerer Bereich und die Werte 6 und 7 als obere Grenze. Das heißt für das Szenario der positiven Arbeitgeberbewertung (MW = 5,27, SD = 1,19), wurden alle ausgeschlossen, die einen Wert schlechter als 3 angegeben haben. Für das Szenario der nicht sozial scheinenden

Abbildung 3: Überblick Experimentalgruppen (eigene Darstellung in Anlehnung an Vredenburg et al. (2020, S.449))

Marke (MW = 1,7, SD = 0,90) wurden keine Bewertungen die größer als 5 waren, akzeptiert. Beim zweiten Manipulationschek wurden alle Personen herausgenommen, die nicht mindestens eine für das Szenario spezifische Tat identifiziert haben. Haben Teilnehmende nur den Aspekt der Solidaritätsbekundung erkannt, der in beiden Szenarien vorhanden war, wurden sie nicht in die Analyse inkludiert. In diesen Fällen konnte nicht identifiziert werden, welches der beiden Szenarien gesehen wurde. Insgesamt wurden 59 Personen ausgeschlossen und die Anzahl der Teilnehmenden von 220 auf 161 reduziert.

Als nächstes wurde geprüft, ob alle Personen mit dem im Markenaktivismus vermittelten Standpunkt übereinstimmen, da dieser Faktor konstant gehalten werden sollte. Die wenigen Teilnehmer, die dies nicht betraf, wurden bereits im vorherigen Schritt ausgeschlossen.

Nach der Bereinigung des Datensatzes wurde er auf die Analyse vorbereitet. Als erstes wurde für die Skepsis-Skala das Item „Ich mag Werbung“ umkodiert. Anschließend wurde die Reliabilität für die Skalen Skepsis gegenüber Werbung, Involvement, Einstellung gegenüber der Marke (vorher und nachher) und wahrgenommene Markenauthentizität geprüft. Alle Skalen wiesen mindestens Ein Cronbachs Alpha größer als 0,7 auf, was für eine aus gute Reliabilität spricht (Nunnally, 1978, S. 245). Darüber hinaus stellte die Faktorenanalyse sicher, dass alle durchschnittlichen extrahierten Varianzen eines Konstrukts über 0,5 lagen, was auf Validität hindeutet (Fornell & Larcker, 1981, S. 45f.). Im nächsten Schritt wurden die Items über die Bildung des Mittelwertes zu einem Konstruktwert verdichtet. Für die wahrgenommene Markenauthentizität wurden zusätzlich Konstruktwerte für die einzelnen Subskalen (Integrität, Glaubwürdigkeit und Symbolismus) gebildet. Anhang 1 zeigt eine Übersicht der Chronbachs Alphas, der Faktorenanalyse und der Mittelwerte. In Tabelle 2 sind zusätzlich die Mittelwerte je nach Experimentalgruppe der Variablen zu sehen, die von der Manipulation betroffen sind. Als letztes wurde die Differenz aus Einstellung nachher und Einstellung vorher errechnet, sodass ein positives Ergebnis für eine Einstellungsverbesserung und ein negatives Ergebnis für eine Verschlechterung steht.

Der finale Datensatz besteht aus 161 Teilnehmenden, von denen 72,05% weiblich, 27,33% männlich sind und 0,62% ihr Geschlecht nicht angeben wollten. Die Mehrheit befindet sich entweder im Studium (45,1%) oder ist berufstätig (42%). Die restlichen 12,9% sind z.B. Schüler (1,2%), in der Ausbildung (3,1%) oder arbeitssuchend (2,5%). Das durchschnittliche Alter ist 27 Jahre und die Alterspanne der Teilnehmenden ging von 16 bis 68. 5% wollten keine Angabe zu ihrem Alter machen. Der größte Anteil, ist zwischen 20-30 (75,2%). Lediglich 5% sind jünger (zwischen 16 und 19 Jahren) und 15,6% sind älter (zwischen 31 und 68 Jahren). Grafische Übersichten der demografischen Daten sind in Anhang 2 zu finden.

Um H1 zu testen, wurde zunächst eine einfaktorielle ANOVA durchgeführt (s. Anhang 3). Die Annahmen der Ausgeglichenheit der Gruppen konnte nicht erfüllt werden. Gruppe 1 (Authentischer Markenaktivismus; n = 38) ist deutlich kleiner als Gruppe 2 (Nicht authentischer Markenaktivismus; n = 123). Ebenso liegt keine Varianzhomogenität vor, da der durchgeführte Levene-Test signifikant ist (p = 0,026) ist. Aus diesem Grund konnte eine gewöhnliche ANOVA nicht betrachtet werden und es wurde ein Welsh-Test durchgeführt. Das dieser signifikant ist (F = 19,124; p < 0,001) bedeutet, dass es signifikante Unterschiede der Gruppen in Bezug auf die Einstellung gibt. Ist der Markenaktivismus authentisch verbessert sich die Einstellung (MW = 0,83). Der Mittelwert von nicht authentischem ist nahe Null (MW = - 0,01) (s. Abb. 4). Da sich authentischer Markenaktivismus positiv auswirkt und nicht authentischer Markenaktivismus im Vergleich zu authentischem schlechter, kann Hypothese 1 bestätigt werden.

Um die weiteren Hypothesen zu testen, wurde der Bootstrappinng Ansatz um Mediationen zu testen von Zhao et al. (2010) sowie Preacher und Hayes (2004) verfolgt. Dafür wurde das „PROCESS“ Macro verwendet und Modell neun ausgewählt. Die Bootstrap-Konfidenzintervalle der indirek-

Tabelle 2: Mittelwerte nach Gruppen (eigene Darstellung)

	Gesamtmittelwerte (n = 161)			
	Authentischer Markenaktivismus (n = 38)	Nicht authentischer Markenaktivismus (n = 123)		
	Gruppe 1 (n = 38)	Gruppe 2 (n = 43)	Gruppe 3 (n = 44)	Gruppe 4 (n = 36)
Einstellung vorher	MW = 3,62; SD = 1,71			
	MW = 4,62; SD = 0,93	MW = 3,31; SD = 1,77		
	MW = 4,62; SD = 0,93	MW = 5,33, SD = 1,03	MW = 2,33, SD = 0,95	MW = 2,10, SD = 0,93
Einstellung nachher	MW= 3,81; SD = 1,82			
	MW = 5,46; SD = 1,13	MW = 3,30; SD = 1,69		
	MW = 5,46; SD = 1,13	MW = 4,32; SD = 1,68	MW = 3,22; SD = 1,46	MW = 2,18; SD = 1,17
Wahrgenommene Markenauthentizität	MW = 3,38, SD = 1,42			
	MW = 4,69; SD = 1,12	MW = 2,97; SD= 1,24		
	MW = 4,69; SD = 1,12	MW = 3,80; SD = 1,31	MW = 2,77; SD = 0,95	MW = 2,23; SD = 0,87
Wahrgenommene Markenauthentizität Dimension: Glaubwürdigkeit	MW = 3,33; SD = 1,50			
	MW = 4,60; SD = 1,16	MW = 2,94; SD = 1,38		
	MW = 4,60; SD = 1,16	MW = 3,75; SD = 1,51	MW = 2,66, SD = 1,09	MW = 2,31; SD = 1,05
Wahrgenommene Markenauthentizität Dimension: Integrität	MW = 3,71; SD = 1,52			
	MW = 5; SD = 1,22	MW = 3,31; SD = 1,38		
	MW = 5; SD = 1,22	MW = 4,16; SD = 1,44	MW = 3,17; SD = 1,07	MW = 2,47; SD = 1,05
Wahrgenommene Markenauthentizität Dimension: Symbolismus	M = 3,08; SD = 1,45			
	MW = 4,45; SD = 1,24	MW = 2,64; SD = 1,24		
	MW = 4,45; SD = 1,24	MW = 3,46, SD = 1,29	MW = 2,45; SD = 0,99	MW = 1,94; SD = 0,88
Authentizität One-Item	MW = 3,29; SD = 1,79			
	MW = 4,95; SD = 1,39	MW = 2,77; SD = 1,57		
	MW = 4,95; SD = 1,39	MW = 3,93; SD = 1,55	MW = 2,45; SD = 1,21	MW = 1,78; SD = 1,07

ten Effekte wurden mit einem Konfidenzniveau von 95% und 5.000 Stichproben geschätzt. Außerdem wurde ein Heteroskedastizität konsistenter Standardfehler- und Kovarianzmatrixschätzer (HC4) verwendet. Authentischer bzw. nicht authentischer Markenaktivismus stellt hier die unabhängige Variable dar, wobei die wahrgenommene Markenauthentizität als Mediator fungiert. Die vorgeschlagenen Moderatoren, die diese Beziehung beeinflussen sollen sind Skepsis sowie Involvement. Die abhängige Variable ist die Einstellung, die über die Einstellungsänderung ermittelt wird.

Zuerst wurde überprüft, ob Skepsis und Involvement eine moderierende Wirkung im Mediationsmodell haben. Es kann weder für Skepsis [-0,112 bis 0,112] noch für Involvement [-0,201 bis 0,038] ein signifikanter Einfluss auf die Beziehung von authentischem Markenaktivismus und wahrgenommener Markenauthentizität festgestellt werden, da das 95-prozentige Konfidenzintervall die Null enthält. Demnach müssen H3 und H4 verworfen werden.

Der Effekt von authentischem Markenaktivismus auf die wahrgenommene Markenauthentizität zeigt jedoch einen signifikanten positiven Einfluss (coeff = 1,660; se(HC4) = 0,232; t = 7,162; p < 0,001). Die Ergebnisse zeigen zudem einen indirekten Effekt von authentischem Markenaktivismus durch die wahrgenommene Markenauthentizität auf die Einstellung zur Marke (effect = 0,395), mit einem 95%igen Konfidenzintervall ohne Null [0,104 bis 0,771]. Da authentischer Markenaktivismus zu einer größeren wahrgenommenen Markenauthentizität führt, kommt es zu einer positiven Einstellungsveränderung. Hypothese H2 kann demnach bestätigt werden. Da auch der Haupteffekt auf einem 10-prozentigen Signifikanzniveau signifikant positiv ist (coeff = 0,435; se(HC4) = 0,260; t = 1,670; p = 0,097) handelt es sich hier um eine komplementäre Mediation (Zhao et al., 2010, S. 199). Eine Übersicht der Ergebnisse der Hypothesentestung ist in Abb. 5 zu sehen.

Im Anschluss an die Hypothesentestung wurden verschiedene Zusatzanalysen durchgeführt. Zunächst wurde die Mediationsanalyse mit weiteren Mediatoren durchgeführt, um die Ergebnisse zu überprüfen und zu vertiefen. Dafür wurden jeweils die einzelnen Authentizitätsdimensionen getestet, so-

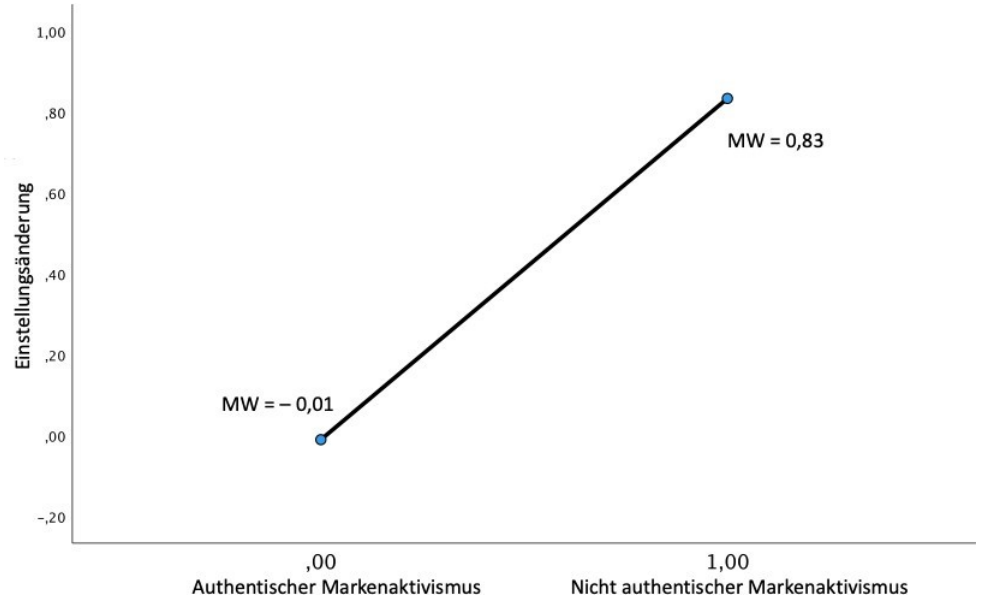


Abbildung 4: Mittelwertunterschiede Authentischer vs. Nicht authentischer Markenaktivismus (eigene Darstellung)

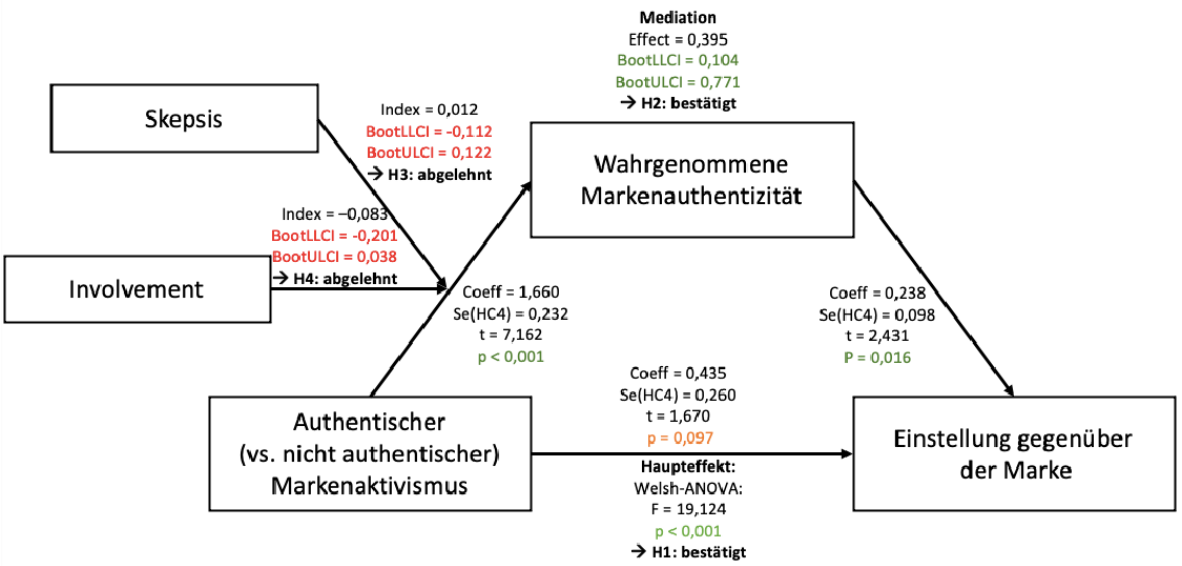
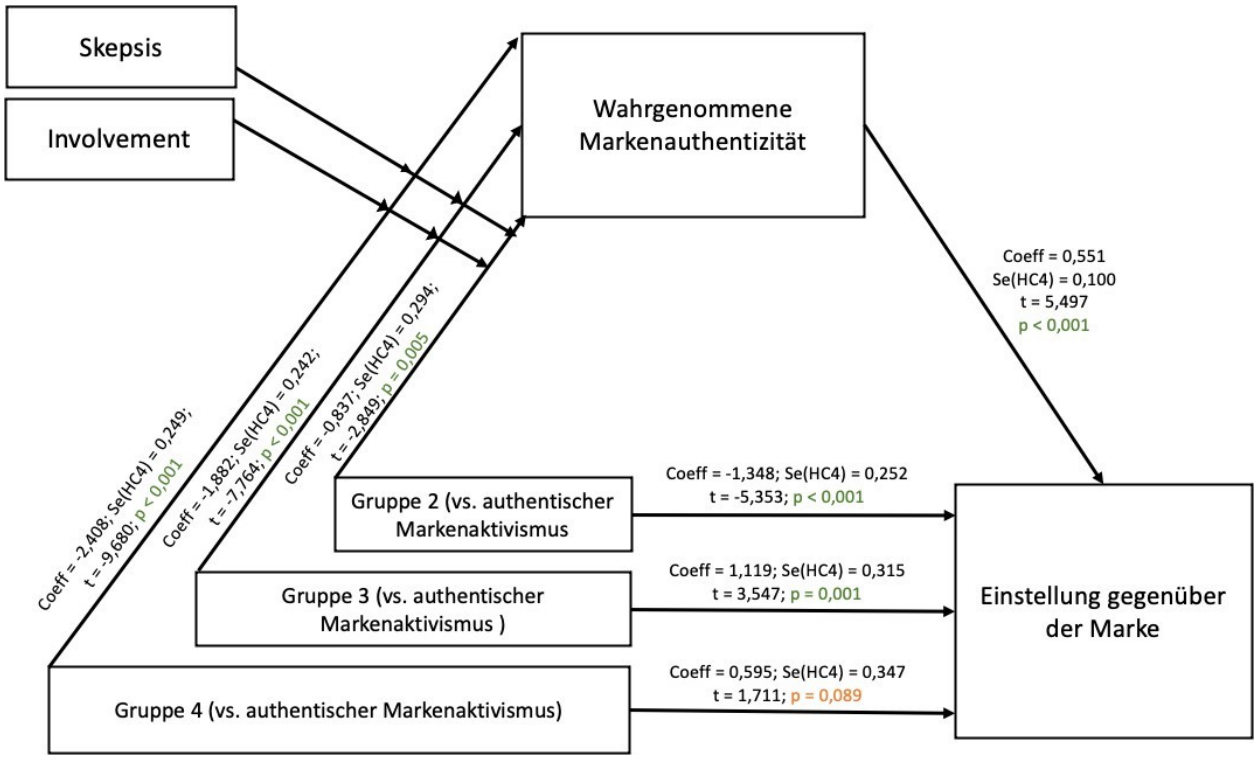


Abbildung 5: Ergebnisse Hypothesentests (eigene Darstellung)

wie die One-Item Messung für Markenauthentizität. Die Ergebnisse weichen nicht wesentlich von den zuvor berichteten Ergebnissen ab, was ein Hinweis auf die Stabilität der Daten und die Robustheit der Theorieentwicklung sein könnte. Die Ergebnisse der Zusatzanalysen sind in Anhang 4 zu finden.

Schließlich wurde eine detaillierte Analyse der verschiedenen Gruppen durchgeführt, die nicht authentischem Markenaktivismus ergeben. Durch die Konzeption des Experiments gibt es insgesamt vier Gruppen, die durch unterschiedliche Kombinationen entweder zu authentischem oder nicht authentischem Markenaktivismus führen. Während authentischer Markenaktivismus nur von einer Gruppe gebildet wird, besteht die Gruppe nicht authentischer Markenaktivismus aus drei Untergruppen (s. Abb. 3). Deshalb besteht der Effekt von nicht authentischem Markenaktivismus aus drei verschiedenen Effekten, deren Einzelbetrachtung sinnvoll ist. Eine Zusammenfassung der Ergebnisse ist in Abb. 6 dargestellt.

Jede Gruppe von nicht authentischem Markenaktivismus wirkt sich im Vergleich zu authentischem Markenaktivismus negativer auf die wahrgenommene Markenauthentizität aus. Dabei hat Gruppe $2_{+/-}$ den geringsten negativen Effekt (coeff $= -0,837$; se(HC4) $= 0,294$; t $= -2,849$; p $= 0,005$), gefolgt von Gruppe $3_{-/+}$(coeff $= -1,882$; se(HC4) $= 0,242$; t $= -7,764$, p $< 0,001$). Gruppe $4_{-/-}$ hat den stärksten negativen Effekt coeff $= -2,408$; se(HC4) $= 0,249$; t $= -9,680$, p $< 0,001$).



Indirekter Effekt der Mediation	**Gruppe → wahrgenommne Authentizität → Einstellung gegenüber der Marke**	
Gruppe 2	Effect = -0,461; BootLLCI = -0,877; BootULCI = -0,136	
Gruppe 3	Effect = -1,037; BootLLCI = -1,518; BootULCI = -0,646	
Gruppe 4	Effect = -1,327; BootLLCI = -1,886; BootULCI = -0,841	
Indizes der partiellen moderieten Mediation:		
	Skepsis	**Involvement**
Gruppe 2	Index = 0,023; BootLLCI = -0,260; BootULCI = 0,337	Index = 0,226; BootLLCI = -0,142; BootULCI = 0,478
Gruppe 3	Index = -0,006; BootLLCI = -0,260; BootULCI = 0,299	Index = 0,083; BootLLCI = -0,198; BootULCI = 0,322
Gruppe 4	Index = -0,043; BootLLCI = -0,299; BootULCI = 0,266	Index = 0,225; BootLLCI = -0,052; BootULCI = 0,448

Abbildung 6: Ergebnisse Detailanalyse 4 Gruppen (eigene Darstellung)

Der indirekte Effekt der Mediation zeigt auf einem 95-prozentigen Konfidenzintervall ohne Null für Gruppe $2_{+/-}$ (effect = −0,461; [−0,877 bis −0,136]), für Gruppe $3_{-/+}$ (effect = −1,037; [−1,518 bis −0,646]) und für Gruppe $4_{-/-}$ (effect = −1,327 [−1,886 bis −0,841)] einen signifikanten negativen Effekt. Durch die negativer wahrgenommene Markenauthentizität wird auch die Einstellung schlechter bewertet. Betrachtet man nun auch den direkten Effekt lassen sich zwei verschiedene Arten von Mediation nach Zhao et al. (2010, S. 199) erkennen. Bei Gruppe $2_{+/-}$ kann ein signifikanter negativer direkter festgestellt werden (coeff = −1,348; se(HC4) = 0,252; t = 5,353; p < 0,001). Es liegt wie in der Hauptanalyse eine komplementäre Mediation vor, da der indirekte und direkte Effekt in dieselbe Richtung wirken. Bei Gruppe $3_{-/+}$ und Gruppe $4_{-/-}$ handelt es sich um konkurrierende Mediationen da jeweils der direkte Effekte signifikant und positiv ist (Gruppe $3_{-/+}$: coeff=1,119; se(HC4) = 0,315; t = 3,547; p = 0,001; Gruppe $4_{-/-}$: coeff = 0,595; se(HC4) =0,347; t = 1,711 p = 0,089), der indirekte Effekt allerdings negativ. Dabei ist der direkte Effekt für Gruppe $4_{-/-}$ lediglich auf einem 10-prozentigen Signifikanzniveau signifikant. Die beiden Effekte wirken demnach in entgegengesetzte Richtungen.

5. Diskussion

5.1. Zusammenfassung und Einordnung der Ergebnisse

Das zentrale Ergebnis dieser Arbeit ist, dass sich authentischer Markenaktivismus positiv auf die Einstellung gegenüber der Marke auswirkt und dass dies maßgeblich von der wahrgenommenen Markenauthentizität abhängt. Durch authentischen Markenaktivismus wird auch die Marke als au-

thentischer angesehen, wodurch es zu einer Einstellungsverbesserung kommt. Im Gegensatz dazu wird bei nicht authentischem Markenaktivismus die Authentizität der Marke negativer wahrgenommen und führt zu einer insgesamt schlechteren Bewertung der Einstellung. Es konnte sowohl die Hypothese zum Haupteffekt (H1) als auch die Hypothese zum Mediationseffekt (H2) bestätigt werden. Diese Erkenntnisse stellen die Antwort auf die erste Forschungsfrage dar, welche wie folgt lautet: *„Wie wirkt sich authentischer (vs. nicht authentischer) Markenaktivismus auf die Einstellung gegenüber der Marke aus?“*. Die Tatsache, dass in der Mediation auch der direkte Effekt signifikant ist und somit eine komplementäre Mediation vorliegt, deutet außerdem darauf hin, dass es neben der wahrgenommenen Markenauthentizität noch andere Aspekte gibt, warum authentischer Markenaktivismus zu einer positiveren Einstellung führt. Einer dieser möglichen Aspekte könnte die durch das soziale Engagement größere wahrgenommene Markenwärme sein. Diese führt laut Stokburger-Sauer et al. (2012, S. 407) ebenfalls zu einer besseren Einstellung aufgrund einer größeren wahrgenommenen Markenidentifikation.

Im Gegensatz zur Haupt- und Mediationshypothese sind die beiden Moderationshypothesen H3 und H4 nicht signifikant. Es kann weder ein Einfluss von Involvement noch von Skepsis auf die Beziehung zwischen authentischem Markenaktivismus und wahrgenommener Authentizität festgestellt werden. Forschungsfrage Nr. 3: *„Lässt sich die Wahrnehmung von authentischem (vs. nicht authentischem) Markenaktivismus durch die Variablen Skepsis oder Involvement beeinflussen?“* kann demnach verneint werden.

Es wurde vermutet, dass Personen, die eine höheres Involvement in das Thema des Markenaktivismus haben, die Informationen detaillierter verarbeiten und vorheriges Wissen einbringen. Dadurch sollte die Authentizität des Markenaktivismus erkannt werden und mit einer höheren Bewertung der wahrgenommenen Markenauthentizität einhergehen. Dies wurde auf Grundlage des Elaboration Likelihood Modells hypostasiert. Dieses bezieht sich im Grunde darauf, wie sich die Einstellung durch persuasive Nachrichten ändert, wie die Informationen verarbeitet werden und wie stabil die resultierende Einstellung ist (Petty & Cacioppo, 1981). Da es konzeptuell näher liegt, wäre es möglich, dass sich das Involvement nicht auf die wahrgenommene Markenauthentizität auswirkt, sondern nur einen Effekt auf die Beziehung von Markenaktivismus und Markeneinstellung hat. In einer zusätzlichen Analyse wurden die Moderatoren auch auf die weiteren Pfade des Modells getestet, zeigen jedoch ebenfalls keine signifikanten Effekte (s. Anhang 5). Daraus lässt sich schließen, dass Involvement (ebenso wie Skepsis) im gesamten konzeptuellen Modell keine Rolle spielt.

Ein möglicher Grund hierfür könnte die Art des betrachteten Involvements sein. Jede Art von Involvement führt zu eigenen idiosynkratischen Ergebnissen, die sich aus dem involviert sein ergeben (Zaichkowsky, 1985, S. 341). Eine Konsequenz von Involvement in eine Kaufentscheidung ist zum Beispiel, dass man mehr Informationen sucht. Involvement für ein Produkt hingegen führt beispielsweise zu einer stärkeren Betrachtung von Produktunterschieden (Zaichkowsky, 1985, S. 341). In dieser Arbeit wurde das Involvement für das Thema Rassismus betrachtet. Es ist möglich, dass nur diese Art von Involvement, also das Involvement für das Thema, nicht zu einer veränderten Wahrnehmung der Authentizität der Marke führt. Stattdessen wäre es möglich, dass, wie in der Arbeit von Fritz et al. (2017), das Involvement in die Marke einen signifikanten Effekt vorweist. Konsumierende, die ein hohes Involvement in die Marke haben, unternehmen wahrscheinlicher kognitive Anstrengungen um eine Marke zu bewerten (Malär et al., 2011, S. 38). Dementsprechend wäre eine Vermutung, dass das Involvement in das Thema dazu führt, dass die Bewertung des Markenaktivismus beeinflusst wird und nicht die Bewertung der Markenauthentizität. Dies müsste jedoch näher untersucht werden.

Skepsis gegenüber Werbung bzw. Marketingaktivitäten sollte sich negativ auf die Beziehung von authentischem Markenaktivismus und wahrgenommener Markenauthentizität auswirken. Es wurde vermutet, dass die Absichten der Marke eher in Frage gestellt werden. Eine mögliche Erklärung für die fehlende Signifikanz könnten die verschiedenen Blickwinkel in situative (Mohr et al., 1998; Skarmeas & Leonidou, 2013) und generelle Skepsis (Boush et al., 1994; Obermiller & Spangenberg, 1998) sein. In diesem Experiment wurde zuerst abgefragt, ob die Teilnehmenden skeptisch gegenüber Werbung sind, und anschließend den verschiedenen Szenarien ausgesetzt. Eine andere Möglichkeit wäre gewesen, die situative Skepsis zu messen, nachdem der Markenaktivismus gezeigt wurde. Die in der Situation ausgelösten Gefühle und Gedanken hätten möglicherweise zu anderen Ergebnissen geführt.

Eine andere Möglichkeit wäre, dass Skepsis aufgrund der Konstruktion des Experiments keinen Effekt hat. Wie eingangs erläutert, kann Skepsis durch ausreichende Beweise entkräftet werden (Mohr et al., 1998, S. 33). Da es sich hier um konstruierte Stimuli handelt, sind alle notwendigen Informationen vorhanden, um sich eine eigene Meinung über die Authentizität der Marke zu bilden. Es wäre möglich, dass Skepsis dadurch in diesem Kontext keine Wirkung zeigt. Außerdem war ein Aspekt, der Skepsis auslöst, die dem Markenaktivismus vorzuwerfende Ursachenausbeutung und die damit verbundenen firmenbezogenen Motive. Das gleiche Argument wie zuvor lässt sich auch hier anwenden. Da in diesem Experiment neutrale Informationen in Form einer Arbeitgeberbewertung oder eines Zeitungsartikels zur Verfügung stehen, könnte es sein, dass keine unternehmensbezogenen Motive für die Marke angenommen werden und daher keine Skepsis aufkommt.

Zuletzt kann auch die zweite Forschungsfrage *„Kann die Konzeptualisierung von authentischem und nicht authentischem Markenaktivismus nach Vredenburg et al. (2020) auch empirisch belegt werden?“,* durch die Ergebnisse der Mediationsanalyse beantwortet werden. Für einen zusätzlichen Erkenntnisgewinn sollte auch die Detailanalyse hinzugezogen werden. Das Markenaktivismus entsprechend der Definition auch als authentisch bzw. nicht authentisch wahrgenommen wird, zeigt sich jeweils an der gesteigerten oder verringer-

ten wahrgenommenen Markenauthentizität. Zusätzlich führt diese höhere bzw. niedrigere Authentizität auch zu negativen Konsequenzen, am Beispiel dieser Arbeit zu einer negativeren Einstellung. Diese beiden Erkenntnisse können als empirischer Beleg für die Konzeptualisierung von authentischem und nicht authentischem Markenaktivismus nach Vredenburg et al. (2020) gesehen werden.

Betrachtet man nun die Detailanalyse der einzelnen Gruppen von nicht authentischem Markenaktivismus, kann dieses Ergebnis grundsätzlich bestätigt werden. Jede der drei Gruppen von nicht authentischem Markenaktivismus wird im Vergleich zu authentischem Markenaktivismus weniger authentisch wahrgenommen. Diese geringere wahrgenommene Markenauthentizität wirkt sich in der Mediation negativ auf die Einstellung zur Marke aus. Allerdings wurde ebenfalls deutlich, dass nicht authentischer Markenaktivismus unter Umständen auch positive Effekte auf die Einstellung gegenüber der Marke haben kann, welche nicht auf die Authentizität zurückzuführen sind.

Nachfolgend werden die Ergebnisse der Detailanalyse genauer betrachtet und diskutiert. Gruppe $4_{-/-}$ wird im Vergleich zu authentischem Markenaktivismus als am wenigsten authentisch angesehen. Dies könnte daran liegen, dass das Engagement nicht zu den Werten der Marke passt und auch die Gestaltung des Markenaktivismus kein ernsthaftes Interesse an der Lösung des gesellschaftlichen Problems erahnen lässt. In Gruppe $2_{+/-}$ wird die Marke als authentischer bewertet als in Gruppe $3_{-/+}$. Eine mögliche Erklärung hierfür könnte sein, dass in Gruppe $2_{+/-}$ durch die positiven Werte eine Art Vertrauensvorschuss vorliegt. Einige Forschungen zeigen, dass wenn eine Marke sozial wirkt, ihr auch positive Eigenschaften in anderen Bereichen zugeschrieben werden (z.B. Chernev und Blair, 2015). Durch die positiven Werte könnte die kritische Bewertung des Markenaktivismus somit milder ausfallen. Es könnte immer noch ehrliches Interesse hinter dem Engagement vermutet werden, welches lediglich schlecht umgesetzt wurde. Insbesondere bei der Konstruktion dieses Experiments, könnte zumindest die Solidaritätsbekundung mit dem Verweis, dass man sich jeher für die Thematik eingesetzt hat als authentisch angesehen werden.

Betrachtet man nun genauer die Mediation, so zeigen die indirekten Effekt, dass jede Gruppe von nicht authentischem Markenaktivismus zu einer geringeren wahrgenommenen Markenauthentizität und dadurch zu einer schlechteren Einstellung führt als bei authentischem Markenaktivismus. Die direkten Effekte zeigen jedoch teilwiese andere Effekte als, die Hauptanalyse zunächst vermuten lässt. Der direkte Effekt von Gruppe $2_{+/-}$ ist komplementär zum indirekten Effekt. Das heißt sowohl direkter als auch indirekter Effekt wirken in dieselbe Richtung und haben einen negativen Einfluss auf die Einstellung. Dies steht im Einklang mit der Theorie. Obwohl die geringere wahrgenommene Authentizität zu einer schlechteren Einstellung führt, liegt bei Gruppe 3-/+ und Gruppe $4_{-/-}$ ein positiver direkter Effekt vor. Hier zeigt sich demnach ein entgegensetzer Effekt. Analog zum authentischen Markenaktivismus könnte auch hier die Einstellungsverbesserung aufgrund einer größeren Identifikation mit der Marke durch das soziale Engagement und der damit einhergehenden größeren wahrgenommenen Markenwärme bestehen. Ob dies tatsächlich der Fall ist, ist vor allem bei Gruppe $4_{-/-}$ fraglich, da unklar ist, inwiefern das täuschende Engagement als warm wahrgenommen werden kann.

Wieso der direkte Effekt bei Gruppe $2_{+/-}$ und Gruppe $4_{-/-}$, bei denen beide opportunistische Hintergedanken deutlich werden, in unterschiedliche Richtungen wirkt, lässt sich mit der Erwartungs-Diskonfirmationstheorie erklären. Aufgrund der positiven Einstellung im Vorfeld besteht bei den Konsumierenden die Erwartung, dass sich die Marke in der Zukunft konform verhält. Es kommt folglich zu einer Enttäuschung, wenn dies nicht der Fall ist (Oliver, 1980, S. 1). In den Szenarien, in denen bereits nach der ersten Manipulation eine negative Einstellung besteht, wird kein sozial verantwortliches Verhalten erwartet. So könnte man im Fall von Gruppe 3-/+ positiv überrascht sein, weil Erwartungen übertroffen wurden, oder im Fall von Gruppe $4_{-/-}$ weder positiv überrascht noch enttäuscht sein, weil das Verhalten der Marke sich nicht aus dem Erwartungsspektrum herausbewegt (Oliver, 1980, S. 2).

5.2. Implikationen für die Theorie

Diese Arbeit leistet einen Beitrag zur aktuellen Forschung über Markenaktivismus und seine Wirksamkeit sowie über die Rolle der Authentizität in diesem Zusammenhang. Die bisherige Forschung hat die Beziehung zwischen Markenaktivismus und Authentizität, z.B. in Form von Woke Washing, oft nur qualitativ untersucht. Es fehlt ein breites Spektrum an quantitativer Forschung in diesem Bereich. Lediglich Hydock et al. (2020) hat die Bedeutung von authentischem Markenaktivismus für Marken mit kleinem Marktanteil empirisch nachgewiesen. Aufgrund der zunehmenden Relevanz von Markenaktivismus sowie Komplexität und Wichtigkeit von Authentizität in diesem Kontext bot sich eine gemeinsame Untersuchung der beiden Aspekte an.

Der aktuelle Forschungsstand kann insbesondere dahingehenden erweitertet werden, dass nicht nur das Übereinstimmen mit dem Standpunkt einer Marke ein relevanter Faktor für den Erfolg oder Misserfolg von Markenaktivismus ist, sondern auch die Authentizität. Authentischer im Gegensatz zu nicht authentischem Markenaktivismus, wirkt sich positiv auf die Einstellung gegenüber der Marke aus. Hier ist die wahrgenommene Markenauthentizität, die durch authentischen Markenaktivismus erhöht wird, ein wesentlicher Treiber für den Effekt.

Darüber hinaus lassen sich empirisch begründete Rückschlüsse auf die Forschung von Vredenburg et al. (2020) ziehen. So können die in dieser Arbeit erlangten Ergebnisse als Belege ihrer Konzeptualisierung von authentischen und nicht authentischen Markenaktivismus gesehen werden. Authentischer (nicht authentischer) Markenaktivismus wird als authentisch (weniger authentisch) angesehen, da er sich positiv (negativ) auf die wahrgenommene Markenauthentizität auswirkt.

Zusätzlich können die Erkenntnisse aus der detaillierten Analyse des nicht authentischen Markenaktivismus erste Hinweise über weitere relevante Effekte darbieten. Die Bewertung der Markenauthentizität sowie die aus dem Markenaktivismus resultierenden Einstellung, können von der vorherigen Einstellung und vermutlich auch durch die damit verbundenen Erwartungen beeinflusst werden.

Als letztes kann die Theorie um die Erkenntnisse erweitert werden, dass die in anderen Arbeiten relevanten Moderatoren Involvement und Skepsis keinen signifikanten Effekt im vorliegenden Kontext von authentischem Markenaktivismus, wahrgenommener Markenauthentizität und der Einstellung gegenüber der Marke haben. Präziser gesagt haben weder Involvement in Bezug auf das Thema des Markenaktivismus noch allgemeine Skepsis gegenüber Werbung bzw. Marketingaktivitäten einen Einfluss auf den Effekt von authentischem Markenaktivismus auf die wahrgenommene Markenauthentizität.

5.3. Implikationen für die Praxis

Da Marken immer mehr in Bedrängnis kommen, in Bezug auf Markenaktivismus zu handeln, stehen Verantwortliche vor der wichtigen Entscheidung, sich aktivistisch zu engagieren oder dies nicht zu tun. In beiden Fällen kann es für Marken sowohl positive als auch negative Folgen haben. Dementsprechend ist es für Praktiker von höchster Priorität, herauszufinden, wie sie negative Folgen vermeiden bzw. von den positiven Folgen profitieren können. Die aus dieser Arbeit erlangten Ergebnisse können dafür genutzt werden.

Eine große Schwierigkeit bei der Betrachtung von Markenaktivismus besteht darin, dass mehrere Komponenten für den Erfolg berücksichtigt werden müssen und es somit auch viele verschiedene Fehlerquellen und Angriffspunkte gibt. Es muss beispielsweise darauf geachtet werden, was der vertretene Standpunkt der Marke auslösen könnte. Im Fall einer Kampagne muss beachtet werden, wie das Thema inhaltlich aufgegriffen und gestaltet wird. Oder, wie es Kern dieser Arbeit ist, ob das Engagement als authentisch wahrgenommen wird.

Um eine klare Empfehlung aussprechen zu können, müssten alle Komponenten gemeinsam betrachtet werden. Selbst wenn sich eine Marke authentisch für die gleichgeschlechtliche Ehe einsetzt, dieses Thema sensibel und gut präsentiert, könnte dies bei Menschen, die dagegen sind, zu negativen Emotionen gegenüber der Marke führen. Deshalb ist es bei den Empfehlungen für die Praxis zu berücksichtigen, dass Authentizität nur eine der zu betrachteten Komponenten ist und man nicht pauschal sagen kann, dass authentischer Markenaktivismus immer zum Erfolg führen wird.

Was die Ergebnisse dieser Studie eindeutig zeigen, ist, dass es für Marken dann lohnenswert sein kann, aktivistisches Engagement zu zeigen, wenn eine Marke von einem Purpose geleitet wird, pro-soziale Werte vertritt und prosoziale Unternehmenspraktiken implementiert hat. Dadurch werden ehrliche Absichten hinter dem Markenaktivismus vermutet. Hierbei ist es dringend notwendig, dass die Art des Aktivismus die sozialen Aspekte und die Problemlösung betont und es nicht den Anschein macht, aus Profitgier und aus eigenem Interesse zu handeln.

Wird der Markenaktivismus demnach als authentisch angesehen, führt es dazu, dass die Marke insgesamt authentischer wahrgenommen wird. Dies ist ein Grund dafür, dass sich die Einstellung zur Marke verbessert. Zusätzlich gibt es weitere nicht näher empirisch untersuchte Aspekte, die sich ebenfalls positiv auf die Einstellung auswirken.

Ein nicht authentisches Engagement sollte grundsätzlich vermieden werden, da es immer zu einer geringeren wahrgenommenen Markenauthentizität führt. Die Ergebnisse zeigen zwar, dass es auch positive Effekte für Marken geben kann, die keine pro-sozialen Werte haben und keinem Purpose folgen, sich aber trotzdem in Markenaktivismus engagieren. Da hier die vorherige Einstellung sehr schlecht ist, wirken mögliche Aspekte wie die Markenwärme eher positiv. So würde sich die Einstellung gegenüber der Marke einerseits verringern, weil die Marke weniger authentisch erscheinen würde. Gleichzeitig könnte sich die Einstellungen aber auch durch andere, empirisch nicht nachgewiesene Aspekte verbessern. Da diese Aspekte nicht näher untersucht wurden, kann nicht empfohlen werden, die Verschlechterung der Authentizität dafür in Kauf zu nehmen. Zudem wurde in dem Experiment nur eine Momentaufnahme betrachtet und es ist nicht bekannt, wie nachhaltig sie sich auf die Markeneinstellung auswirken.

5.4. Limitationen und zukünftige Forschung

Bei der Betrachtung dieser Arbeit müssen einige Limitation beachtet werden. Diese beziehen sich zum einen auf die Gestaltung des Experiments und den damit verbundenen Einschränkungen der Gültigkeit der Ergebnisse. Zum anderen gibt es inhaltliche und theoriebasierte Limitationen, die vor allem bei der Interpretation zu berücksichtigen sind. Letztere bieten zahlreiche Möglichkeiten für zukünftige Forschungen.

Aufgrund der geplanten Auswertung einer Haupt- und einer Detailanalyse ergeben sich in der Hauptanalyse sehr unterschiedliche Gruppengrößen. In ihr sollen die Gruppen authentischer Markenaktivismus ($n = 38$) und nicht authentischer Markenaktivismus ($n = 123$) untersucht werden. Nicht authentischer Markenaktivismus besteht dabei aus drei verschiedenen Gruppen, die in der Detailanalyse näher untersucht werden. Darüber hinaus wurden im Rahmen der Datenbereinigung 61 Personen aus der Analyse ausgeschlossen, was einem Anteil an allen Befragten von ca. 25% entspricht. Somit sind die Gruppen nicht nur unausgeglichen, sondern es handelt sich insgesamt um eine eher kleine Stichprobe.

Darüber hinaus sollten Einschränkungen hinsichtlich der Repräsentativität der Stichprobe berücksichtigt werden. Um sicherzustellen, dass die Ergebnisse nicht auf die Beschaffenheit der Probe zurückzuführen sind, sollte das Experiment mit anderen Stichproben wiederholt werden.

Um eine Einstellungsänderung zu messen, wurde das Experiment so angelegt, dass die Marke nach der ersten Manipulation nicht ausschließlich positiv oder negativ erscheint.

So sollte vermieden werden, dass die Endpunkte der Skala bereits bei der ersten Bewertung der Einstellung erreicht

werden. In einigen Fällen wurde jedoch nach der ersten Manipulation die schlechteste Einstellungsbewertung abgegeben. Dadurch wurde eine mögliche Einstellungsveränderung nach der zweiten Manipulation nicht mehr messbar. So kann in diesen Fällen keine aussagekräftige Einstellungsänderung erfasst werden. Dies ist insbesondere für Gruppe $4_{-/-}$ von Bedeutung, da sowohl bei der ersten als auch bei der zweiten Manipulation eine negative Bewertung erwartet wurde.

Für Experimente üblich ist die externe Validität eine zu betrachtende Limitation. Es wurde versucht, ein möglichst realistisches Szenario darzustellen, indem Elemente wie eine Arbeitgeberbewertung und ein Zeitungsartikel verwendet wurden. Trotzdem handelt es sich um ein konstruiertes Szenario, dessen Ergebnisse nur bedingt auf die Realität übertragbar sind. In diesem Zusammenhang ist es beispielweise fraglich, wie stabil und interpretierbar die Einstellung gegenüber der Marke ist, da sie in diesem Experiment lediglich auf Basis von zwei Informationen gebildet wurde. Je nachdem, wie man Einstellungen betrachtet, entweder als stabile Struktur im Gedächtnis (z.B. Fishbein und Ajzen, 1975) oder als spontan konstruiertes Urteil (z.B. Conrey und Smith, 2007; Schwarz, 2007), könnte unterschiedlich argumentiert werden.

Die verwendeten Skalen wurden alle aus der Literatur entnommen und bei Bedarf auf das Thema angepasst. Im Falle der Skala zur wahrgenommenen Markenauthentizität wurde die Dimension Kontinuität nicht berücksichtigt. Somit ist das vollständige Modell nach Morhart et al. (2015) nicht abgebildet. Auch dies ist ein Bereich für mögliche weitere Untersuchungen. Zur Replikation der hier gewonnen Ergebnisse könnten auch die Definitionen anderer Forschenden für wahrgenommene Markenauthentizität verwendet und daraufhin untersucht werden, ob sie zu denselben Schlussfolgerungen kommen.

Wie bereits bei den Implikationen erwähnt, besteht die für die Praxis am relevantesten theoretische Einschränkung darin, dass hier nur das Konzept der Authentizität isoliert betrachtet wurde. Dies war das Ziel dieser Arbeit und wurde bewusst so gewählt. Für aussagekräftige und praktisch relevante Implikationen sollte man jedoch zusätzlich das Zusammenspiel mit anderen Faktoren betrachten. Zukünftige Forschung könnte sich also nicht nur auf den Einfluss der Authentizität allein konzentrieren. Es könnte zusätzlich geprüft werden, wie das Zusammenspiel zwischen Authentizität und dem von der Marke vertretenen Standpunkt oder anderen für den Markenaktivismus relevanten Aspekten ist.

Neben einer gesamtheitlichen Betrachtung von Markenaktivismus bietet diese Arbeit zusätzlich eine Grundlage, um speziell den Aspekt der Authentizität und die Forschung darüber im Kontext von Markenaktivismus auszuweiten. Es konnte gezeigt werden, dass der Effekt von authentischem Markenaktivismus auf die Einstellung von wahrgenommener Markenauthentizität mediiert wird.. Außerdem gab es zusätzlich einen signifikanten direkten Effekt in der Mediation. Dieser lässt darauf schließen, dass neben der Authentizität des authentischen Markenaktivismus auch andere Aspekte einen positiven Einfluss auf die Einstellung haben. Die Theorie legt nahe, dass zum Beispiel das allgemeine pro-soziale Verhalten der Marke, ein solcher Aspekt sein könnte. Weitere Untersuchungen könnten dies experimentell belegen oder zusätzliche Faktoren identifizieren. Darüber hinaus erwiesen sich die Moderatoren als nicht signifikant. Weitere Untersuchungen könnten durchgeführt werden, um herauszufinden, was die Wahrnehmung der Markenauthentizität beeinflusst. Darüber hinaus könnte sich die Suche nach Moderatoren auch auf andere Beziehungen ausweiten, wie die zwischen authentischem Markenaktivismus und der Einstellung gegenüber der Marke. Da authentischer Markenaktivismus ein wertebasiertes Konstrukt ist, wäre die Untersuchung verschiedener Wertetypen als Moderatoren eine interessante Option.

Für den Bereich der Unternehmenskommunikation stellen sich die folgenden Bereiche als besonders interessant dar. Zum einen könnte die Zugang zu den übermittelten Informationen untersucht werden. Die hier zugrundeliegende Definition impliziert, dass ausreichende Kenntnisse über eine Marke vorhanden sein müssen, um die Authentizität bewerten zu können. In der Realität verfügen die Menschen möglicherweise nicht über diese Informationen, insbesondere bei unbekannten Marken. Es könnte demnach untersucht werden, wie die Bewertung der Authentizität variiert, wenn die Menschen viele, wenige oder gar keine Informationen über eine Marke haben. Außerdem wurde in diesem Experiment ein Zeitungsartikel und eine Arbeitgeberbewertung als Informationsmedium verwendet. Es könnte weiter untersucht werden, ob auch die Quelle der Informationen einen Einfluss auf die Wahrnehmung der Authentizität des Markenaktivismus hat (vgl. Du et al., 2010; Mukherjee und Althuizen, 2020).

Weiterhin könnte untersucht werden, welche aktivistischen Handlungen der Marke besonders positiv und authentisch und welche als weniger positiv und weniger authentisch wahrgenommen werden. Zusätzlich könnte getestet werden, welchen Einfluss ein einmaliges oder wiederkehrendes Engagement auf die Wahrnehmung der Authentizität hat.

In dieser Arbeit wurde authentischer Markenaktivismus besonders aus einer wertebasierten Perspektive betrachtet. Wie eingangs erwähnt, ist ein weiterer Bereich, der im Zusammenhang mit Markenaktivismus, untersucht wird, das Ausmaß inwieweit ein Engagement thematisch zur Marke passt (z.B. Champlin et al., 2019; Nan und Heo, 2007). Ist es sinnvoll, dass sich eine Marke, die für Luxus und Prestige steht, für die Bekämpfung der Armut einsetzt, oder dass sich eine Marke, die für Männlichkeit wirbt, für die Rechte der Frauen kämpft? Diese Sicht stellt eine interessante Erweiterung der vorliegenden Forschung dar. Es könnten insbesondere konzeptionelle Ähnlichkeiten und Unterschiede herausgearbeitet werden.

Als letzter Punkt und direkte Erweiterung dieser Arbeit sollte genauer untersucht werden, wie die vorherige Einstellung gegenüber einer Marke die Wahrnehmung von Authentizität und Einstellungsänderungen beeinflusst. Dieser Aspekt wurde in dieser Arbeit lediglich angerissen, da das Hauptaugenmerk auf den Gesamteffekt von authentischem und nicht authentischem Markenaktivismus lag. Die De-

tailanalyse zeigte jedoch unterschiedliche Effekte auf die Einstellung gegenüber der Marke und die wahrgenommen Markenauthentizität in Abhängigkeit der Zusammensetzung der Gruppen für nicht authentischen Markenaktivismus. Dies war vor allem darauf zurückzuführen, dass die Gruppen unterschiedliche Ausgangspunkte hatten (gute vs. schlechte Einstellung). Da Menschen in der Realität bereits bestehende Einstellung zu den ihnen bekannten Marken haben, könnte eine solche Erweiterung der Forschung besonders für die Praxis von großem Interesse sein.

Literatur

Aaker, J. L. (1997). Dimensions of Brand Personality. *Journal of Marketing Research*, *34*(3), 347–356. https://doi.org/10.1177/002224379703400304

Accenture. (2019). To affinity and beyond: From me to we, the rise of the purpose-led brand. Verfügbar 7. September 2022 unter https://www.accenture.com/_acnmedia/thought-leadership-assets/pdf/accenture-competitiveagility-gcpr-pov.pdf

Ajzen, I., & Fishbein, M. (1977). Attitude-Behavior Relations: A Theoretical Analysis and Review of Empirical Research. *Psychological Bulletin*, *84*(5), 888–918. https://doi.org/10.1037/0033-2909.84.5.888

Baldinger, A. L., & Rubinson, J. L. (1996). Brand Loyalty: The Link Between Attitude and Behavior. *Journal of Advertising Research*, *36*, 22–34.

Beverland, M. B. (2005). Crafting Brand Authenticity: The Case of Luxury Wines*. *Journal of Management Studies*, *42*(5), 1003–1029. https://doi.org/10.1111/j.1467-6486.2005.00530.x

Boush, D. M., Friestad, M., & Rose, G. M. (1994). Adolescent Skepticism toward TV Advertising and Knowledge of Advertiser Tactics. *Journal of Consumer Research*, *21*(1), 165–175. https://doi.org/10.1086/209390

Brown, S., Kozinets, R. V., & Sherry, J. F. (2003). Teaching Old Brands New Tricks: Retro Branding and the Revival of Brand Meaning. *Journal of Marketing*, *67*(3), 19–33. https://doi.org/10.1509/jmkg.67.3.19.18657

Brown, S. P., & Stayman, D. M. (1992). Antecedents and consequences of attitude toward the ad: A meta-analysis. *Journal of Consumer Research*, *19*(1), 34–51.

Bruhn, M., Schoenmüller, V., Schäfer, D., & Heinrich, D. (2012). Brand authenticity: Towards a deeper understanding of its conceptualization and measurement. *Advances in Consumer Research*, *40*, 567–576.

Bücker, T. (2022). So helfen Unternehmen der Ukraine. Verfügbar 10. September 2022 unter https://www.tagesschau.de/wirtschaft/unternehmen/ukraine-hilfen-firmen-101.html

Campbell, M. C., & Kirmani, A. (2000). Consumers' use of persuasion knowledge: The effects of accessibility and cognitive capacity on perceptions of an influence agent. *Journal of Consumer Research*, *27*(1), 69–83. https://doi.org/10.1086/314309

Carp, S. (2018). NFL Confirms Long-Term Extension with Nike. *Sports Pro Media*. Verfügbar 10. September 2022 unter https://www.sportspromedia.com/news/nfl-nike-uniform-deal/

Champlin, S., Sterbenk, Y., Windels, K., & Poteet, M. (2019). How brand-cause fit shapes real world advertising messages: a qualitative exploration of 'femvertising'. *International Journal of Advertising*, *38*(8), 1240–1263. https://doi.org/10.1080/02650487.2019.1615294

Changing Markets Foundation. (2022). Does Primark Care. Verfügbar 10. September 2022 unter https://greenwash.com/brands/primark/

Chernev, A., & Blair, S. (2015). Doing Well by Doing Good: The Benevolent Halo of Corporate Social Responsibility. *Journal of Consumer Research*, *41*(6), 1412–1425. https://doi.org/10.1086/680089

Cinelli, M. D., & LeBoeuf, R. A. (2020). Keeping It Real: How Perceived Brand Authenticity Affects Product Perceptions. *Journal of Consumer Psychology*, *30*(1), 40–59. https://doi.org/10.1002/jcpy.1123

Collins, C. M., Steg, L., & Koning, M. A. S. (2007). Customers' Values, Beliefs on Sustainable Corporate Performance, and Buying Behavior. *Psychology and Marketing*, *24*(6), 555–577. https://doi.org/10.1002/mar.20173

Conrey, F. R., & Smith, E. R. (2007). Attitude Representation: Attitudes as Patterns in a Distributed, Connectionist Representational System. *Social Cognition*, *25*(5), 718–735. https://doi.org/10.1521/soco.2007.25.5.718

Dodd, M. D., & Supa, D. (2015). Testing the Viability of Corporate Social Advocacy as a Predictor of Purchase Intention. *Communication Research Reports*, *32*(4), 287–293. https://doi.org/10.1080/08824096.2015.1089853

Dodd, M. D., & Supa, D. W. (2014). Conceptualizing and Measuring "Corporate Social Advocacy" Communication: Examining the Impact on Corporate Financial. *Public Relations Journal*, *8*(3), 23.

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

EDEKA ZENTRALE Stiftung & Co. KG. (2022a). Freiheit ist ein Lebensmittel. *LinkedIn*. Verfügbar 11. September 2022 unter https://www.linkedin.com/posts/edeka-ag_wir-alle-verfolgen-in-diesen-stunden-mit-activity-6902656587727556608-hnau

EDEKA ZENTRALE Stiftung & Co. KG. (2022b). Hilfe für die Ukraine #StandingWithUkraine. *LinkedIn*. Verfügbar 11. September 2022 unter https://www.linkedin.com/posts/edeka-ag_standwithukraine-activity-6905522240755245056-A-UC

Edelman. (2019). 2019 Edelman Trust Barometer Special Report: In Brands We Trust? *Edelman Trust Barometer Annual Global Study*. Verfügbar 10. September 2022 unter https://www.edelman.com/sites/g/files/aatuss191/files/2019-06/2019_edelman_trust_barometer_special_report_in_brands_we_trust.pdf

EU-Kommission. (2011). Eine neue EU-Strategie (2011-14) für die soziale Verantwortung der Unternehmen (CSR), Mitteilung der Kommission an das Europäische Parlament, den Rat, den euröpäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen, KOM(2011) 681, Brüssel.

Ewing, D. R., Allen, C. T., & Ewing, R. L. (2012). Authenticity as meaning validation: An empirical investigation of iconic and indexical cues in a context of "green" products. *Journal of Consumer Behaviour*, *11*(5), 381–390. https://doi.org/10.1002/cb.1382

Faircloth, J. B., Capella, L. M., & Alford, B. L. (2001). The Effect of Brand Attitude and Brand Image on Brand Equity. *Journal of Marketing Theory and Practice*, *9*(3), 61–75. https://doi.org/10.1080/10696679.2001.11501897

Fishbein, M., & Ajzen, I. (1975). *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*.

Forehand, M. R., & Grier, S. (2003). When Is Honesty the Best Policy? The Effect of Stated Company Intent on Consumer Skepticism. *Journal of Consumer Psychology*, *13*(3), 349–356. https://doi.org/10.1207/S15327663JCP1303_15

Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, *18*(1), 39–50.

Fritz, K., Schoenmueller, V., & Bruhn, M. (2017). Authenticity in Branding – Exploring Antecedents and Consequences of Brand Authenticity. *European Journal of Marketing*, *51*(2), 324–348. https://doi.org/10.1108/EJM-10-2014-0633

Gaski, J. F., & Etzel, M. J. (1986). The Index of Consumer Sentiment toward Marketing. *Journal of Marketing*, *50*(3), 71–81. https://doi.org/10.1177/002224298605000306

Grayson, K., & Martinec, R. (2004). Consumer Perceptions of Iconicity and Indexicality and Their Influence on Assessments of Authentic Market Offerings. *Journal of Consumer Research*, *31*(2), 296–312. https://doi.org/10.1086/422109

Huffington Post. (2011). "Hubby Hubby": Ben & Jerry's Shows Gay Marriage Support With New Flavor. Verfügbar 10. September 2022 unter https://www.huffpost.com/entry/hubby-hubby-ben-jerrys-sh_n_273872

Hydock, C., Paharia, N., & Blair, S. (2020). Should Your Brand Pick a Side? How Market Share Determines the Impact of Corporate Politi-

cal Advocacy. *Journal of Marketing Research*, *57*(6), 1135–1151. https://doi.org/10.1177/0022243720947682

Iqbal, N. (2019). Woke washing? How brands like Gillette turn profits by creating a conscience. *The Guardian*. Verfügbar 10. September 2022 unter https://www.theguardian.com/media/2019/jan/19/gillette-ad-campaign-woke-advertising-salving-consciences

Kanter, D. L., & Wortzel, L. H. (1985). Cynicism and alienation as marketing considerations: Some new ways to approach the female consumer. *Journal of Consumer Marketing*, *2*(1), 5–15. https://doi.org/10.1108/eb038816

Key, T. M., Keel, A. L., & Olson, E. M. (2021). Brand activism change agents: Strategic storytelling for impact and authenticity. *Journal of Strategic Marketing*, 1–17. https://doi.org/10.1080/0965254X.2021.1904435

Kotler, P., & Sarkar, C. (2017). „Finally, Brand Activism!“ *The Marketing Journal*. Verfügbar 9. September 2022 unter https://www.marketingjournal.org/finally-brand-activism-philip-kotler-and-christian-sarkar/

Krugman, H. E. (1967). The Measurement of Advertising Involvement. *Public Opinion Quarterly*, *30*(4), 583–596. https://doi.org/10.1086/267457

Kununu. (2022). Bewerte deinen Arbeitgeber. Verfügbar 11. September 2022 unter https://www.kununu.com/de/insights?forms=bewerten,gehalt,kultur

Laurent, G., & Kapferer, J.-N. (1985). Measuring Consumer Involvement Profiles. *Journal of Marketing Research*, *22*(1), 41–53. https://doi.org/10.1177/002224378502200104

MacKenzie, S. B., Lutz, R. J., & Belch, G. E. (1986). The Role of Attitude Toward the Ad as a Mediator of Advertising Effectiveness: A Test of Competing Explanations. *Journal of Marketing Research*, *23*, 130–143. https://doi.org/10.1177/002224378602300205

Malär, L., Krohmer, H., Hoyer, W. D., & Nyffenegger, B. (2011). Emotional Brand Attachment and Brand Personality: The Relative Importance of the Actual and the Ideal Self. *Journal of Marketing*, *17*(4), 35–52.

Marketing Science Institute. (2016). Research priorities 2016–2018. Verfügbar 10. September 2022 unter http://www.msi.org/uploads/articles/MSI_RP16-18.pdf

Marquis, C., Toffel, M. W., & Zhou, Y. (2016). Scrutiny, Norms, and Selective Disclosure: A Global Study of Greenwashing. *Organization Science*, *27*(2), 483–504. https://doi.org/10.1287/orsc.2015.1039

McDonnell, M.-H. (2016). Radical Repertoires: The Incidence and Impact of Corporate-Sponsored Social Activism. *Organization Science*, *27*(1), 53–71. https://doi.org/10.1287/orsc.2015.1017

Mirzaei, A., Wilkie, D. C., & Siuki, H. (2022). Woke brand activism authenticity or the lack of it. *Journal of Business Research*, *139*, 1–12. https://doi.org/10.1016/j.jbusres.2021.09.044

Mishra, D. P., Heide, J. B., & Cort, S. G. (1998). Information Asymmetry and Levels of Agency Relationships. *Journal of Marketing Research*, *35*(3), 277–295. https://doi.org/10.2307/3152028

Mitchell, A. A., & Olson, J. C. (1981). Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude? *Journal of Marketing Research*, *18*(3), 318–332. https://doi.org/10.1177/002224378101800306

Mohr, L. A., Eroğlu, D., & Ellen, P. S. (1998). The Development and Testing of a Measure of Skepticism Toward Environmental Claims in Marketers' Communications. *Journal of Consumer Affairs*, *32*(1), 30–55. https://doi.org/10.1111/j.1745-6606.1998.tb00399.x

Morhart, F., Malär, L., Guèvremont, A., Girardin, F., & Grohmann, B. (2015). Brand authenticity: An integrative framework and measurement scale. *Journal of Consumer Psychology*, *25*(2), 200–218. https://doi.org/10.1016/j.jcps.2014.11.006

Mukherjee, S., & Althuizen, N. (2020). Brand activism: Does courting controversy help or hurt a brand? *International Journal of Research in Marketing*, *37*(4), 772–778. https://doi.org/10.1016/j.ijresmar.2020.02.008

Nan, X., & Heo, K. (2007). Consumer Responses to Corporate Social Responsibility (CSR) Initiatives: Examining the Role of Brand-Cause Fit in Cause-Related Marketing. *Journal of Advertising*, *36*(2), 63–74. https://doi.org/10.2753/JOA0091-3367360204

Napoli, J., Dickinson, S. J., Beverland, M. B., & Farrelly, F. (2014). Measuring consumer-based brand authenticity. *Journal of Business Research*, *67*(6), 1090–1098. https://doi.org/10.1016/j.jbusres.2013.06.001

Norman, J. (2012). Chick-fil-A's controversy fallout. Verfügbar 10. September 2022 unter https://www.ocregister.com/2012/08/12/chick-fil-as-controversy-fallout/

Nunnally, J. C. (1978). *Psychometric Theory* (2. Aufl.).

Obermiller, C., Spangenberg, E., & MacLachlan, D. L. (2005). Ad skepticism: The Consequences of Disbelief. *Journal of Advertising*, *34*(3), 7–17. https://doi.org/10.1080/00913367.2005.10639199

Obermiller, C., & Spangenberg, E. R. (1998). Development of a Scale to Measure Consumer Skepticism Toward Advertising. *Journal of Consumer Psychology*, *7*(2), 159–186. https://doi.org/10.1207/s15327663jcp0702_03

Oliver, R. L. (1980). A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions. *Journal of Marketing Research*, *17*(4), 460–469.

Peloza, J., & Shang, J. (2011). How can corporate social responsibility activities create value for stakeholders? A systematic review. *Journal of the Academy of Marketing Science*, *39*(1), 117–135. https://doi.org/10.1007/s11747-010-0213-6

Petty, R. E., & Cacioppo, J. T. (1981). Issue Involvement As a Moderator of the Effects on Attitude of Advertising Content and Context. *NA - Advances in Consumer Research*, *8*, 20–24.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.

Pew Research Centre. (2017). In First Month, Views of Trump Are Already Strongly Felt, Deeply Polarized. Verfügbar 10. September 2022 unter https://www.pewresearch.org/politics/wp-content/uploads/sites/4/2017/02/02-16-17-Political-release.pdf

Pirsch, J., Gupta, S., & Grau, S. L. (2007). A Framework for Understanding Corporate Social Responsibility Programs as a Continuum: An Exploratory Study. *Journal of Business Ethics*, *70*(2), 125–140. https://doi.org/10.1007/s10551-006-9100-y

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*, 717–731. https://doi.org/10.3758/BF03206553

Reed, A., Forehand, M. R., Puntoni, S., & Warlop, L. (2012). Identity-based consumer behavior. *International Journal of Research in Marketing*, *29*(4), 310–321. https://doi.org/10.1016/j.ijresmar.2012.08.002

Reuters. (2017). Starbucks responds to criticism over refugee hiring blowback. https://www.reuters.com/article/us-usa-trump-immigration-starbucks-idUSKBN15H24F

Schwarz, N. (2007). Attitude Construction: Evaluation in Context. *Social Cognition*, *25*(5), 638–656. https://doi.org/10.1521/soco.2007.25.5.638

Sherif, C. W., Kelly, M., Rodgers, J., H. L., Sarup, G., & Tittler, B. I. (1973). Personal involvement, social judgment, and action. *Journal of Personality and Social Psychology*, *27*(3), 311–328. https://doi.org/10.1037/h0034948

Shetty, A. S., Venkataramaiah, N. B., & Anand, K. (2019). Brand activism and millennials: An empirical investigation into the perception of millennials towards brand activism. *Problems and Perspectives in Management*, *17*(4), 163–175. https://doi.org/10.21511/ppm.17(4).2019.14

Skarmeas, D., & Leonidou, C. N. (2013). When consumers doubt, Watch out! The role of CSR skepticism. *Journal of Business Research*, *66*(10), 1831–1838. https://doi.org/10.1016/j.jbusres.2013.02.004

Sobande, F. (2019). Woke-washing: "Intersectional" femvertising and branding "woke" bravery. *European Journal of Marketing*, *54*(11), 2723–2745. https://doi.org/10.1108/EJM-02-2019-0134

Spears, N., & Singh, S. N. (2004). Measuring Attitude toward the Brand and Purchase Intentions. *Journal of Current Issues & Research in Advertising*, *26*(2), 53–66. https://doi.org/10.1080/10641734.2004.10505164

Steiner, C. J., & Reisinger, Y. (2006). Understanding Existential Authenticity. *Annals of Tourism Research*, *33*(2), 299–318. https://doi.org/10.1016/j.annals.2005.08.002

Stokburger-Sauer, N., Ratneshwar, S., & Sen, S. (2012). Drivers of Consumer–Brand Identification. *International Journal of Research in Marketing*, *29*(4), 406–418. https://doi.org/10.1016/j.ijresmar.2012.06.001

Tagesschau. (2022). McDonald's verkauft russische Filialen. Verfügbar 10. September 2022 unter https://www.tagesschau.de/wirtschaft/unternehmen/mcdonalds-russland-rueckzug-101.html

The Guardian. (2019). Nike's 'Dream Crazy' Advert Starring Colin Kaepernick Wins Emmy. https://www.theguardian.com/sport/2019/sep/16/nikes-dream-crazy-advert-starring-colin-kaepernick-wins-emmy

Trilling, L. (1972). *Sincerity and Authenticity*. Harvard University Press.

von Wangenheim, F., & Bayón, T. (2007). The Chain from Customer Satisfaction via Word-of-Mouth Referrals to New Customer Acquisition. *Journal of the Academy of Marketing Science*, *35*(2), 233–249. https://doi.org/10.1007/s11747-007-0037-1

Vredenburg, J., Kapitan, S., Spry, A., & Kemper, J. A. (2020). Brands Taking a Stand: Authentic Brand Activism or Woke Washing? *Journal of Public Policy*, *39*(4), 444–460. https://doi.org/10.1177/0743915620947359

Wang, N. (1999). Rethinking Authenticity in Tourism Experience. *Annals of Tourism Research*, *26*(2), 349–370. https://doi.org/10.1016/S0160-7383(98)00103-0

Wettstein, F., & Baur, D. (2016). "Why Should We Care about Marriage Equality?": Political Advocacy as a Part of Corporate Responsibility. *Journal of Business Ethics*, *138*(2), 199–213. https://doi.org/10.1007/s10551-015-2631-3

White, K., Argo, J. J., & Sengupta, J. (2012). Dissociative versus Associative Responses to Social Identity Threat: The Role of Consumer Self-Construal. *Journal of Consumer Research*, *39*(4), 704–719. https://doi.org/10.1086/664977

Zaichkowsky, J. L. (1985). Measuring the Involvement Construct. *Journal of Consumer Research*, *12*(3), 341–352.

Zaichkowsky, J. L. (1986). Conceptualizing Involvement. *Journal of Advertising*, *15*(2), 4–34. https://doi.org/10.1080/00913367.1986.10672999

Zhao, X., Lynch, J. G., & Chen, Q. (2010). Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. *Journal of Consumer Research*, *37*(2), 197–206. https://doi.org/10.1086/651257

Junior Management Science 9(3) (2024) 1634-1664

Junior Management Science

www.jums.academy
ISSN: 2942-1861

The Employees' Entrepreneurial Mindset: The Influence of Perceived Supervisor Effort on the Employees' Entrepreneurial Passion

Athanasios Konstantinos Kallinikidis

Technical University of Munich

Abstract

This paper examines the effect of perceived supervisor effort on the employees' entrepreneurial passion. The study combines theories on emotional contagion, goal contagion, and self-regulation to explain the underlying mechanisms for occurring phenomena. Case-based research delivered the data to investigate the relationship between perceived supervisor effort and the employees' entrepreneurial passion. The data revealed that proximity to the founders, entrepreneurial-relatedness of the employee's tasks, and initial entrepreneurial passion work as antecedents of the combined mechanism of contagion and self-regulation. The findings suggest that employees are affected positively by the perception of high effort and negatively by the perception of low effort in their passion for entrepreneurship when proximity to the founders, entrepreneurial-relatedness of the tasks, and initial entrepreneurial passion are high. However, the perception of high effort can decrease entrepreneurial passion when employees were initially low passionate about entrepreneurship. This work contributes to the literature by providing a theoretical model that describes how perceived supervisor effort impacts entrepreneurial passion on an employee-level outcome.

Keywords: contagion; employee; entrepreneurial effort; entrepreneurial passion; perceived effort

1. Introduction

Entrepreneurial passion is metaphorical "the fire of desire" (Cardon et al., 2009, p. 515) that drives efforts in the context of new venture creation and is therefore essential to research in the field of entrepreneurship (Gielnik et al., 2015). Founders with a high entrepreneurial passion are more creative, motivated, and successful (Cardon et al., 2005, 2013; Chen et al., 2009). Effort is another relevant construct in entrepreneurship as it is considered a driver for successful launches of new businesses (Foo et al., 2009). Scholars insist that effort reflects the purpose of mobilizing resources and energy to achieve the desired objective (Dik & Aarts, 2007). They interpret effort as a signal to pursue goals. Consequently, passion and effort are tremendously substantial constructs for successful entrepreneurship.

There is much research on how passion influences entrepreneurial effort (Baum et al., 2001; Cardon et al., 2009). While the direction of influence seems intuitively correct to start from emotion like passion leading to behavior like effort (Russell, 2003), Gielnik et al. (2015) found a causal relationship for the other direction: entrepreneurial effort driving entrepreneurial passion. Breugst et al. (2012) paved the way for a new literature stream on entrepreneurial passion. They put the entrepreneur's employee into the spotlight and researched the effect of perceived entrepreneurial passion on the employees' commitment toward the nascent venture. They found significant relationships between passion and affective commitment.

Moreover, Hubner et al. (2020) researched the effects of perceiving passion. They contend that entrepreneurs can develop in their employees a sense of passion for their tasks, especially if that passion did not exist previously, and enhance performance by expressing passion. However, the work of Breugst et al. (2012) and Hubner et al. (2020) has shown that there is little knowledge of the effects of the employees' perception of their supervisor in the entrepreneurial context. This research stream, combined with the novel finding of Gielnik et al. (2015) that effort is a factor stimulating

DOI: https://doi.org/10.5282/jums/v9i3pp1634-1664

entrepreneurial passion, reveals a research gap on how employees' emotions are affected by the perceptions of their supervisor's behavior. Therefore, the question arises whether and how the conclusion of Gielnik et al. (2015) holds for the framework provided by scholars like Breugst et al. (2012) or Hubner et al. (2020) with putting the employee in the focus of the analysis.

The theoretical framework of this study merges the theoretical approaches of the respective research streams of Breugst et al. (2012) and Gielnik et al. (2015). Similar to the mentioned academic contributions, I will draw on the emotional contagion theory (Hatfield et al., 1993; Platow et al., 2005) to illustrate the transfer of emotions between two individuals. Besides, I use the self-regulation theory (Bandura et al., 1999; Carver & Scheier, 1982; Locke & Latham, 2002) to describe the within-person relationship of effort and passion in the direction that behavior influences emotion. First, relying on emotional contagion theory explains that displayed positive affect by the entrepreneur can generate affective reactions in the employee. Following this theory, affect can be transferred from entrepreneur to employee during social interactions at work, influencing the employee's emotions. Therefore, perceiving entrepreneurial passion can make employees more passionate about entrepreneurship. Second, applying self-regulation theories explains that effort can trigger positive affect when employees make progress. Progress reduces the discrepancy to the desired goal, which leads to experiencing positive emotions (Carver, 2006). This mechanism indicates that employees can become more passionate about entrepreneurship when they put effort into entrepreneurial tasks. To build the bridge between those theoretical frameworks, I will add the theory on goal contagion (Dik & Aarts, 2008; Palomares, 2013) to point out how perceived effort can trigger the employee's effort. Goal contagion builds on the priming mechanism (Laurin, 2016). It explains that the employee must perceive an external stimulus that activates a mental representation which he or she then accesses later and adapts as his or her own.

Breugst et al. (2012) and Hubner et al. (2020) asked how the perception of entrepreneurial passion can influence employee-level outcomes. Gielnik et al. (2015) questioned the unique direction of passion driving effort while investigating how the reverse effect could occur. As a result of both approaches, I aim to combine both streams and research the effect of perceived supervisor effort on the employee's entrepreneurial passion. This results in the following research question:

> *How does the employees' perception of their supervisor's effort influence the employees' entrepreneurial passion?*

I adopted a qualitative methodology to discover an answer to my research question. This approach is effective in research settings with at least limited theory or knowledge about how a process operates, such as in the current research (Lee et al., 1999). I used a multiple case study approach that helps to develop new insights into a research subject that has not yet been thoroughly investigated (Eisenhardt, 1989). A case study technique is a suitable methodology because there is only limited theory on how perceived supervisor effort influences the employees' entrepreneurial passion. A multiple case study approach helps answer research questions that start with "how" or "why" (Edmondson & McManus, 2007). The unit of analysis I want to investigate is the perception and emotions of current or former employees of small ventures as they are or were in frequent and direct contact with the entrepreneurs (Breugst et al., 2012). Therefore, the best-suited sample for data collection is the actual entrepreneurial employee. I used theoretical sampling to select the appropriate population for this study (Miles & Huberman, 1994). I video-interviewed eight entrepreneurial employees and used a semi-structured setting that provided flexibility in data collection (Edwards & Holland, 2013). To strengthen the validity of this study, I relied on the triangulation logic (Jick, 1979) and enriched the cases with secondary data.

The study makes several contributions to the literature as the interplay of entrepreneurial passion and effort from the employees' point of view tackles the fields of emotion, behavior, and entrepreneurship. I contribute by using the work of Gielnik et al. (2015) within the framework scholars like Breugst et al. (2012) and Hubner et al. (2020) provided by investigating whether perceived effort can drive passion while focusing on the entrepreneurial employee. My approach results in novel findings for the merged research streams. Besides, this study enriches the research on factors driving entrepreneurial passion as asked for by previous contributions (Cardon et al., 2012; Gielnik et al., 2015; Murnieks et al., 2014). Furthermore, this study contributes to research on leadership as an entrepreneurial task which has not been explored much (Breugst et al., 2012) by showing how the perceived leader behavior can impact the emotions of his employees.

Overall, this paper extends the literature by providing a theoretical model that describes how perceived supervisor effort impacts the employees' entrepreneurial passion and derives testable propositions. The findings suggest that employees' entrepreneurial passion can be affected by the mere perception of supervisor effort. The perception of the supervisor's effort can have a noticeable effect on the employee, depending on factors such as initial passion, proximity to the founders, and the entrepreneurial-relatedness of the employees' tasks. Then, employees perceiving high effort are more likely to put up more effort on their own, which in turn stimulates an increase in entrepreneurial passion. On the other hand, employees who perceive low effort may exert less effort themselves, which leads to a decline in entrepreneurial passion. Additionally, if an employee initially felt less passionate about entrepreneurship, too much effort may cause that passion for waning. Another important finding of this study is that perceptions of strong passion can outweigh perceptions of poor effort, suggesting that perceptions of emotion may significantly impact employees' emotions more than perceptions of behavior.

For practitioners, my findings help current employees and entrepreneurs understand the outcome of entrepreneurial passion with perceived effort as an input factor. Then, entrepreneurs can work on their communication skills and show their efforts the right way to achieve changes in the entrepreneurial passion of their employees. As many firms want their employees to have an entrepreneurial mindset, developing the employees' entrepreneurial passion is highly beneficial. Employees with high entrepreneurial passion could improve innovation, develop inventions, or start a venture inside the firm.

2. Theory

This contribution aims to investigate the effects of the perception of an entrepreneur's effort on the employees' entrepreneurial passion. Current research provides knowledge that paves the way for three different possible paths. Taken together, they can explain how the perception of entrepreneurial effort can affect the entrepreneurial passion experienced by the employees. Two paths are interpersonal as they describe how the transfer of passion or effort from an entrepreneur to an employee occurs. The third path is intrapersonal and describes how passion develops through effort.

Literature found answers to how passion transfers from one person to another via perception and its effects (Breugst et al., 2012; Brundin et al., 2008; Cardon, 2008; Hubner et al., 2020) (Path I). Moreover, another stream of literature focused on how effort and goals transfer between people through perception (Aarts et al., 2004; Breugst et al., 2020; Dik & Aarts, 2007, 2008; Loersch et al., 2008; Palomares, 2013) (Path II). Finally, scholars found out about the counter-intuitive intrapersonal path from effort influencing passion in entrepreneurship (Gielnik et al., 2015; Lex et al., 2020) (Path III). They diverge from the thought that emotions drive behavior (Russell, 2003) but converge on emotions as a feedback system (Baumeister et al., 2007). This system allows us to account for emotion as the output of behavior. Combining these literature streams will bring us closer to answering the research question. In this study, the first path, which describes a transfer of emotion, works without additional considerations. However, the second path, which describes a transfer of behavior, and the third path, which describes how behavior stimulates emotion, act as a unit in this framework.

First, I will define and describe the core concepts of this work: entrepreneurial passion, effort, and the entrepreneurial employee. As this work studies entrepreneurial passion, entrepreneurial effort, and the entrepreneurial employee, it is vital to have a common understanding of them and look at the established knowledge of these concepts and subjects in the literature. Second, I will review the literature on the three paths starting from entrepreneurial effort leading to the employees' entrepreneurial passion and show the current state of research. Third, I will explain the relevant theories for the different paths to understand how perceived effort can develop into entrepreneurial passion.

2.1. Entrepreneurial effort, passion, and the employee

2.1.1. Entrepreneurial passion

Scholars agree on the importance of passion in the entrepreneurial context (Cardon et al., 2013; Gielnik et al., 2015; Lex et al., 2019; Murnieks et al., 2014). However, the definition of passion varies across academics. While Baum and Locke (2004) define passion as the love for one's work, other scholars emphasize the attraction to engage in certain activities in their definitions (Philippe et al., 2010; Vallerand et al., 2003). Literature concurs on the emotional dimension of passion (Chen et al., 2009) and defines entrepreneurial passion as intense and positive emotions felt during entrepreneurial tasks (Cardon et al., 2009; Drnovsek et al., 2016; Hubner et al., 2020). As entrepreneurial passion is affective by nature (Cardon et al., 2009), this paper looks at it as a phenomenon of experience during certain activities than as a trait of an entrepreneur, which is in line with current literature (Cardon et al., 2013; Vallerand et al., 2003). For this academic contribution, I decompose entrepreneurial passion as a strong, positive, and affective emotion toward entrepreneurial tasks. These emotions are not exclusive to the entrepreneur but can be experienced by employees too, which is then called employee passion response (Hubner et al., 2020).

The literature distinguishes between three types of entrepreneurial passion: passion for inventing, passion for founding, and passion for developing (Breugst et al., 2012; Cardon et al., 2009, 2013). Passion for inventing includes identifying and exploring the market for new business opportunities to develop and invent new products, services, and prototypes. Passion for founding arises when passion occurs in creating new ventures to exploit specific opportunities. Expanding and extending a business after its creation is part of the passion for developing.

Entrepreneurial passion is related to different characteristics. For instance, entrepreneurial passion contributes to the success of entrepreneurs (Breugst et al., 2012), precisely their success in starting and managing a business (Cardon et al., 2009), and to the growth and success of startups (Baum & Locke, 2004; Cardon et al., 2017; Drnovsek et al., 2016). Besides, scholars describe passion as a driver for motivation (Cardon et al., 2005; Chen et al., 2009) while having effects on creativity (Baron, 2008; Cardon, 2008) and commitment (Breugst et al., 2012). Furthermore, entrepreneurial passion is connected to evaluations, fundraising (Mitteness et al., 2012), and recruiting essential employees (Cardon, 2008). However, the most relevant characteristic of passion for this paper is that passion is contagious, which is one core assumption in entrepreneurship literature (e.g., Hubner et al. 2020). It means that employees can perceive and catch the passion of the entrepreneur. Emotional contagion theory explains this phenomenon.

2.1.2. Entrepreneurial effort

People say that if someone wants to achieve something, they must put effort into it. Nowadays, people speak of effort as the engine to achieve goals. According to Dik and

Aarts (2007), effort mirrors the intention to mobilize energy and resources to accomplish the desired goal. They see effort as a signal for motivational goal pursuit. In entrepreneurship, this means that people inside an organization work intensively on entrepreneurial tasks to pursue and attain specific desired goals (Foo et al., 2009). For this paper, I define entrepreneurial effort concurrently to literature as intensive work in entrepreneurial tasks to reach entrepreneurial goals. Entrepreneurs and employees can perform entrepreneurial tasks, which is standard practice in startups nowadays.

Effort is often connected to success and confidence. The rationale is that the more effort someone puts into a task, the higher the likelihood of succeeding (Bandura et al., 1999; Lex et al., 2019). If people believe in their skills and abilities to succeed in entrepreneurial tasks, they are more likely to intensify their work and thus be successful (Gatewood et al., 2002). Entrepreneurial tasks embody goals to give the entrepreneurial effort the necessary direction and clarity (Cardon et al., 2009). People who put effort into a particular task to achieve a goal express how important their goals are to them and how much they value them (Corcoran et al., 2020; Dik & Aarts, 2008). Similarly to passion, effort is contagious (Breugst et al., 2020), meaning that, in the entrepreneurial context, effort is transferable from the entrepreneur to the employee. Goal contagion theory, supported by theory on social motivation, can explain this mechanism.

2.1.3. Entrepreneurial employee

As the operating environment for firms becomes more complex and dynamic, given fast and discontinuous change, the whole organization has to act entrepreneurially (Hitt, 2000). Although employees play a different role compared to the entrepreneur as they have not found the firm, they can also be involved and put effort into entrepreneurial tasks and, therefore, experience passion while engaging them. Employees in new ventures take on entrepreneurial tasks regularly by themselves or collaborate closely with the entrepreneur when working on entrepreneurial processes, for example, during entrepreneurial opportunity development or when contributing with their innovative ideas (Hubner et al., 2020). A relevant research stream tries to understand the phenomena occurring when employees act entrepreneurially inside a firm (e.g., Moriano et al. 2014). Intrapreneurship supports the idea of employees participating in entrepreneurial tasks.

Consistent with previous research, this paper highlights the role of the employee and his perception of the entrepreneur in the context of research on entrepreneurial passion (e.g., Breugst et al. 2012; Brundin et al. 2008; Hubner et al. 2020) and therefore aligns with the call from Cardon (2008) to extend the literature in that area. Researching what employees perceive of their supervisors is essential as this perception may implicate their behavior (Dik & Aarts, 2007). Brundin et al. (2008) argued for the importance of employees who think and act as entrepreneurs as they create new knowledge, products, and services and discover valuable business opportunities

In nascent ventures, employees perceive a close relation to their supervisors and hence feel responsible for and highly involved in entrepreneurial activities while they develop a passion for them (Breugst et al., 2012). As passionate entrepreneurs are crucial to venture success and employees are also vital for the performance of new ventures (e.g., Hayton 2003), Cardon (2008) deducted that passionate employees are beneficial. Furthermore, employees can also benefit from the positive characteristics of having entrepreneurial passion, so passionate employees may be more creative or motivated than employees with less passion who, in addition, are less successful at work tasks (Ho & Pollack, 2014).

2.2. Paths from perceived entrepreneurial effort to employee entrepreneurial passion

2.2.1. Contagion of entrepreneurial passion (Path I)

Cardon (2008) delivered the starting point in research on the contagion of entrepreneurial passion to employees and stated it as “a new area of inquiry in the field of entrepreneurship” (Cardon, 2008, p. 84). She summarized the literature’s opinion on the importance of passion for entrepreneurial success. Cardon further argued that having passionate employees is also relevant and deducted how entrepreneurs could transfer their passion to increase the employees’ passion. She argued that entrepreneurs display their passion and used emotional contagion theory to describe the process of passion transfer.

Brundin et al. (2008) conducted empirical research to explore the connection between the supervisor’s emotional display and the employees’ willingness to act entrepreneurially. Although passion was not explicitly stated and measured as an emotion in that paper, it fits into their category of positive emotions. Satisfaction also finds a place, as positive emotions are typical for passion. For their experiment, they collected data from 91 employees from 31 different small Swedish firms who were in frequent contact with the CEO. Overall, they claimed that displaying positive emotions from the entrepreneur would “put employees in positive moods with respect to their entrepreneurial motivation” (Brundin et al., 2008, p. 238).

Breugst et al. (2012) extended the work of Brundin et al. (2008), participated in the research stream of entrepreneurial passion, and investigated the impacts of perceived entrepreneurial passion on the employees’ commitment to new ventures. They analyzed how the perception of entrepreneurial passion for inventing, developing, and founding affects employee commitment. To explain the phenomenon, they used emotional contagion and goal-setting theory. They argued that perceived passion leads to the experience of positive affect at work and affects goal clarity. Then, the constructs of positive affect at work and goal clarity affect employee commitment. The result showed a positive relationship between passion for inventing and developing with venture commitment, while passion for founding is negatively associated with that construct. They explain this negative association as entrepreneurs who are passionate

about founding may signal their employees that they might leave the firm to start another one again. Besides, they claimed that the affective path showed a higher magnitude than the path via goal clarity through their data which is in line with other scholars who highlight the affective nature of passion (e.g., Cardon 2008). They measured these effects by conducting a quantitative study through a survey, receiving responses from employees from over 100 early-stage ventures. They also found out that the venture stage does not influence the impact of perceived entrepreneurial passion.

Hubner et al. (2020) enriched the literature on the contagion of entrepreneurial passion and its effect on employees. Again, scholars used emotional contagion theory to explain the mechanism of entrepreneurial passion transfer from the entrepreneur to the employee. They were the first to empirically demonstrate the contagion mechanism of entrepreneurial passion using two complementary studies. In the first field study, they matched cross-sectional survey data from German employees with their supervisors, where they received over 200 responses. In the second study, conducted through an experimental design, they hired 321 freelance workers from an online platform and told them that a real entrepreneur hired them. Then, they watched video messages that contained the manipulation for entrepreneurial passion where they got their tasks explained. Their findings suggested that the contagion of passion exists, leading the employees to put extra effort into their tasks and enhance other work-related outcomes. They reasoned that a higher employee passion response makes employees invest more energy into accomplishing entrepreneurial tasks as they connect these tasks to positive emotions, which then results in higher commitment towards the tasks (Breugst et al., 2012; Visser et al., 2013). According to them, entrepreneurs can make their employees passionate about entrepreneurial tasks, especially when they have not been passionate beforehand, and also stimulate their work performance (e.g., effort).

To summarize, entrepreneurship research acknowledges the importance of entrepreneurial passion and its contagious nature. Scholars found out that passion is transferrable between people. In entrepreneurship, passion can transfer from the entrepreneur to the employee. Furthermore, the literature also stated that the perception of entrepreneurial passion or related emotions might also trigger subsequent beneficial behavior, for example, commitment or the willingness to act entrepreneurially. Therefore, employees may perceive their supervisors' entrepreneurial passion and become passionate about entrepreneurial tasks. Emotional contagion theory describes the working mechanism of this phenomenon, which I will explain later.

2.2.2. Contagion of goals and the role of perceived effort (Path II)

Literature on goal contagion assumes that people connect other people's goals to their behavior. Aarts et al. (2004) investigated how and under which circumstances goal contagion occurs. They tested for goal contagion in six different experimental designs in their contribution. They exposed their participants to written scenarios of goals on a computer screen, for example, to make money. Afterward, in a behavioral setting, they tested for the goal contagion of the participants. In that setting, the participants could strive for the goal. However, it required a different behavior than that provided in the written scenario, so they could rule out that the participants mimicked the presented actions. They used Dutch undergraduate students as a randomly assigned control or goal condition group sample. Overall, they provided the foundation for other scholars to build on goal contagion theory as they found strong support in their data for their hypothesis that people unconsciously take on the goals implied by other people. They claimed that people might become more similar in what they want and hence what they plan for the future when goal contagion occurs. Furthermore, they found out that goals pursued under unacceptable ways or conditions or with improper manners become less desirable, and goal contagion becomes less likely to occur.

Dik and Aarts (2007, p. 728) advanced this idea by highlighting the role of perceived effort in goal contagion as a "basic characteristic of motivational goal-directed behavior" that facilitates the occurrence of goal contagion. They see perceived effort as a cue to other people's goals. This cue helps the perceiver to account for the goal-directed behavior and thus helps to discover the specific goals that motivate the acting agent. With the help of a self-produced video that implied the goal of helping, they built an experimental study to test for the impact of perceived effort on goal contagion. The sample for this study consisted of overall 116 Dutch undergraduate students. The video showed a large ball that tried to help another smaller ball to free a stuck kite out of a tree. The ball had to search for a ladder inside a room with four doors. The number of doors manipulated the ball's effort to open to access the ladder. Then, they exposed their participants to a word completion task where they had to come up with the word help to check if the goal of helping was accessed successfully. Additionally, the manipulation check was conducted by asking the participants how much effort the ball had put into searching for the ladder. They later conducted a similar second study, replacing the word completion task with a lexical decision task to ensure that goal inference occurred spontaneously without the participants' conscious awareness. During a third experiment, they wanted to find evidence of changes in actual behavior after watching the animated films. After being exposed to the same videos as in the previous experiments, they asked their participants if they wanted to fill out another questionnaire without getting a reward for their participation. They could freely decide if they wanted to leave the laboratory or if they wanted to volunteer without being asked to help directly. The data on this experiment showed that perceiving more effort led the participants to a stronger pursuit of the goal. All in all, their findings propose a linear relationship between effort and goal inferences. The more effort is shown, the more accessible the goal for the perceiver of this effortful behavior.

In a subsequent study, Dik and Aarts (2008) investigated whether the perception of another peoples effort to attain a goal that might be yet unknown to the perceiver triggers the perceiver's motivation to find out about the goal. They assumed that an effortful behavior signalizes the goal as valuable to the pursuer. To test their hypothesis, they used three experiments where Dutch undergraduate students observed an agent who pursued an unrevealed goal with either low, medium, or high effort. Afterward, they tested the participants' motivation to infer the unknown goal. The two experiments used a text comprehension task and an animated film to demonstrate an agent's effortful behavior and let them self-capture their motivation for goal inference. In the third experiment, they used a clicking task on a computer connected to the animated film of the previous experiment as a behavioral measure to account for goal inference motivation. Thereby, they extended their study by adding actual behavior in a spontaneous manner. They found evidence for the perceived effort to be connected to goal inference: "people become more motivated to find out the goal of an actor's behavior whenever this behavior is characterized by more effort." (Dik & Aarts, 2008, p. 750). They further argue that goal inference is essential in collaborative tasks and that other people's goals are a relevant source of environmental information. For them, the motivation to infer goals is the basis for goal contagion, which is enhanced by the perception of effortful behavior.

In the context of goal contagion, Loersch et al. (2008) investigated the understanding of the role of group belonging. They hypothesized that it is more likely for goal contagion to occur when the actor and the perceiver belong to the same group. In their study, they showed their participants self-created videos of people playing racquetball with either competitive or cooperative behavior. In the competitive version, the actors played more intensely, while in the cooperative version, they played more slowly with less intention to win the game and instead kept the ball up in the air. They labeled the videos so the participants could categorize the actors as joint group members. They did this in the shared group membership version by showing an overlay in the video with the text of the respective university of the participants who were students. Afterward, they measured the goal activation by asking the participants to imagine themselves being a coach at an American football team and devise a strategy to win a game. The researchers categorized the provided strategies by competitiveness. They found a significant relationship between goal contagion and group membership. Those who viewed competitive behavior by members of their group wanted to implement a more competitive strategy than the others, whereas there was no difference in the non-membership group.

Palomares (2013) further validated research on goal contagion as he was the first to study authentic conversations after the effect's examination in written scenarios or videos. He argued for its importance because goal contagion "[is] highly social and rooted in interaction" (Palomares, 2013, p. 76). He let undergraduate students in the US form conversations with each other where one had to perceive what the other had to strive for a goal. The pursuer received a three-level goal to test for goal specificity. Then, the perceiver filled out a questionnaire to test for goal contagion. Overall, he claimed his study to be a successful support and replication of the study of Aarts et al. (2004) but in a more natural and realistic setting.

Breugst et al. (2020) investigated the contagion of effort in new venture management teams, taking the research into the entrepreneurship context. They based their work on social motivation theory, suggesting that effort is contagious. However, they tested the boundaries of social motivation theory to see under which circumstances contagion is hampered or facilitated. In a longitudinal study, they collected data through surveys from 161 cofounders of 64 different teams managing early-stage ventures. In their research, they found out that when a teammate puts effort into the startup, it plays an essential role in triggering the effort of the focal manager. Although they did not find support for contagion to happen automatically, they found effort contagious at the management level of new venture teams when threats emerge. These threats are a low performance of the own venture and environmental hostility.

To conclude, research acknowledges that people put effort into tasks to achieve specific goals, and these goals, as well as their efforts, can be transferred from one person to another. As these effects are more robust if the people belong to the same group, this idea also applies to the entrepreneur and his employee. Literature found evidence that people can infer the goals of others, find them valuable and start to pursue them by themselves and, thus, put effort into tasks. If they already perceive higher levels of effort from the agent, these goals are more likely to transfer. Goal contagion theory explains this phenomenon's working mechanism, which I will present later. In addition to the role of effort in making goal contagion more likely, effort is also contagious. Although scholars investigated this with new venture management teams as the unit of analysis, the arguments presented also hold for the context of the employee-entrepreneur relationship as they often collaborate closely in small ventures. Therefore, perceiving entrepreneurial effort can make the employee increase his or her effort on entrepreneurial tasks through goal and effort contagion.

2.2.3. Impact of effort on entrepreneurial passion (Path III)

When talking about feelings and behavior, it seems intuitive that behavior follows emotion. We often say that people do certain things because they feel a certain way. When Russell (2003) theorized about the psychological construction of emotion, he argued that emotional states influence behavior. Foo et al. (2009) studied how feelings influence effort in the entrepreneurial context. They focused on the affect-effort link because entrepreneurship is an affective process and effort is a significant factor for new venture success. They argued, using self-regulation theory, that negative affect is a sign of slower progress, and thus entrepreneurs will increase subsequent effort. Positive affect, on the other side,

signalizes that things are going well and opens the scope of attention. They hypothesized that positive affect leads entrepreneurs to focus more on the future, which results in them engaging with extra effort in venture tasks beyond what is required immediately. They tested their hypothesis using an experience sampling methodology with a sample of 46 entrepreneurs who had to fill out short surveys on their phones multiple times a week. Overall, they found support in the data for their claims.

Baumeister et al. (2007) contradict the predominant view by arguing that emotions occur in feedback loops where they can emerge as descendants of people's behavior. They assume that all psychological processes (e.g., emotion) exist to influence behavior partly but not directly. They theorized that emotion influences behavior through a feedback system. They used the example of guilt to explain their theory. In the example, they describe a person who causes distress to a friend, so he or she feels guilty afterward. Because of the experienced guilt, the person thinks about what he or she has done wrong to avoid a similar feeling in the future. If there is a similar situation next time, the person may adapt his or her behavior so that it does not cause distress. So, first, there was the behavior, then there was the emotion which resulted in a change in later behavior to avoid this emotion. They argue that emotion as feedback is helpful to modify behavior and, therefore, also valuable for goal pursuit as behavior directs toward a goal. The behavior will be adjusted in goal pursuit to experience more positive emotions as they signalize progress toward a cherished goal.

Gielnik et al. (2015) followed this view by claiming that it is not only emotion that drives behavior, but it is also behavior that drives emotion. Notably, they investigated how entrepreneurial effort influences entrepreneurial passion and found evidence for effort to predict changes in passion. They draw upon theories of self-regulation, that is, control, goal-setting, and social cognitive theory, to explain the underlying mechanisms. They conducted a field study over eight weeks to find support for their claims. Their sample consisted of 54 German entrepreneurs who had to complete an online survey weekly where they should report their work-related effort and passion. Additionally, they run a laboratory experiment to further investigate the causal chain from effort to passion. Therefore, they took undergraduate students who first completed a questionnaire to capture entrepreneurial passion and the commitment to invest effort. Afterward, they received the task of developing a business idea into a more mature business plan. After completing this task, they had to complete a second and third survey that served as manipulation checks and outcome variables as entrepreneurial passion was measured. To manipulate effort, they varied the working time for the task. Overall, their findings indicate that entrepreneurial effort predicts changes in entrepreneurial passion. New venture progress as a mediator of this effect and free choice as a moderator of the mediated effect provides the underlying causal link.

Other scholars like Lex et al. (2020) based their work on this idea. They accepted effort as an antecedent of entrepreneurial passion while using self-regulation theories to cover new models on the development of passion in entrepreneurship over time. They posit that passion develops in a feedback loop dependent on entrepreneurial self-efficacy and performance. Then, this performance gets cognitively evaluated. To test their model, they used 65 entrepreneurs from Tanzania across their first study's three phases of the entrepreneurial process. They collected data through structured interviews in person as wells as through questionnaires after the interview and a subsequent questionnaire later in the process. Data collection resulted in three different measurements for each participant during the study. Their first study provided evidence of entrepreneurial performance's impact on positive feelings. Their second study aimed to extend those findings. There, they offered an entrepreneurship training program over 12 weeks to simulate all stages and significant tasks of entrepreneurship. The sample consisted of 150 Tanzanian students, and data was collected through questionnaires by the end of each training week. Their results replicated the findings from the first study successfully. They extended them by providing evidence for the influence of entrepreneurial performance on identity centrality, another antecedent of passion, over the more extended 12-week period. They conducted a third study to generalize and extend the present findings to provide evidence for the mediation effect of entrepreneurial self-efficacy between the effect of entrepreneurial performance on positive feelings. They used a similar methodology to the second study but adapted the time frame and the questionnaires to account for the mediation effect. They found support in the collected data for their hypothesis. Overall, they found support for their recursive and reciprocal model of the development of passion over time. According to their model, passion develops due to evaluating one's performance, with entrepreneurial self-efficacy mediating this relationship. They would answer the question of what came first, passion or performance, the following: " [...] passion and performance develop jointly and iteratively over time in a circular manner, not necessarily with a starting point inherent to unidirectional relationships." (Lex et al., 2020, p. 26).

All in all, literature has differing views on what came first, the emotion or the behavior. One stream of the literature assumes that entrepreneurial passion influences entrepreneurial effort, while the other stream makes the assumption of the opposite direction valid. The view on entrepreneurial passion as part of a feedback system brings both views together and lets them exist concurrently. Therefore, entrepreneurial passion may influence entrepreneurial effort, but it may also be true that entrepreneurial passion arises through the mere exertion of effort on entrepreneurial tasks. For the employee, this means that while working on entrepreneurial tasks, he may become passionate about those tasks. Theories on self-regulation, namely control, goal-setting, and social cognition theory, can explain this phenomenon.

2.3. Explaining theories: how and why the three paths work

2.3.1. Contagion of goals and emotions

I use theories of emotional and goal contagion to explain two different ways in which the perceived behavior of the entrepreneur effects the employee passion response. Theories on contagion will explain the interpersonal effects from the entrepreneur to the employee. Although both theories are not the same, they describe how either the emotion of passion or the behavior of effort is transferable from one individual to another. Literature defines contagion as "a process in which a person or group influences the emotions or behavior of another person or group through the conscious or unconscious induction of emotion states and behavioral attitudes" (Schoenewolf, 1990, p. 50). Contagion, therefore, describes that perceived emotions and behaviors of others can influence the own emotions and behavior.

I use emotional contagion theory to explain how the perception of an entrepreneur's passion at work triggers an employee's passion response. I use goal contagion to describe how effort is carried from one person to another through working on goals.

Goal contagion

To understand goal contagion, suppose that the employee and the entrepreneur of a small new venture sit in the same office and close to each other. They frequently interact during the workday and can engage with each other often. The employee and the supervisor work closely together on the same tasks to achieve specific goals to help the firm in making process and succeed. During these close engagements, the entrepreneur will often set the goals or formulate them with the employee in collaboration. Due to the close collaboration, the employee will perceive that the entrepreneur puts effort into the stated goals and will see himself in the same boat as his supervisor. According to goal contagion theory, the employee may see the goals now as his own and will work intensively on the tasks to achieve them. Therefore, effort transfers from the entrepreneur to the employee. Concurrently, everyday life shows us the phenomenon: we get inspired by other people, especially if they have a higher position. When we observe how inspiring people reach their goals, we sometimes try to set similar goals and develop similar behavior to achieve them. When the younger brother perceives that the older brother is writing good grades at school while spending much time studying, the younger brother will likely try to spend more time in preparation to write better grades.

Aarts et al. (2004, p. 24) summarized the definition of the term goal from literature and described it as "a mental representation of the desired state that may pertain to a behavior [...] or an outcome [...]". An example of the behavior can be entrepreneurial effort, and an example of the outcome can be the success of the entrepreneurial firm. They further describe the process of goal contagion as "automatic adoption and pursuit of goals that others are perceived to strive for" (Aarts et al., 2004, p. 24). Automatic means that the process starts without the need for consciousness or intention of the employee. According to them, the perception of behavior triggers goal contagion. Furthermore, Dik and Aarts (2008) supported that idea by theorizing that goals become desirable for a person by observing others working hard on their achievement. Empirical findings of other scholars give additional evidence for this idea (e.g., Corcoran et al. 2020).

The working mechanism of goal contagion is that of a priming phenomenon that needs three conditions to occur (Laurin, 2016). First, the employee must perceive an external stimulus that activates a mental representation. An example of an external stimulus in our case can be that the employee sees his entrepreneur spending much time on product innovation, so the employee will infer that the entrepreneur's goal is to bring a new product to the market to increase firm performance. The external stimulus is the entrepreneur, and the mental representation is the goal. Second, the goal has to remain accessible for a particular duration after its activation, which depends on the motivational relevance (Eitam & Higgins, 2010). In our example, the employee has to perceive the entrepreneur working on the mentioned tasks frequently, and he has to care to some degree about the goals. As the employee is in the same firm as his supervisor, he has a personal interest in the firm performing well. Third, the perceiver must misattribute the accessibility of the mental representation as own desire (Loersch & Payne, 2014). When the employee later tries to investigate how he can contribute to firm performance by doing product innovation, he may misattribute that this was his supervisor's goal and assume that he now wants to pursue his own firm performance goal.

Perceived effort is a catalyst for goal contagion to occur. The literature stated that it is more likely for someone to infer the goal of the other if one sees the other putting more effort into a task (Dik & Aarts, 2007, 2008; Palomares, 2013). The reason is that the perceiver will interpret the goal as of higher value due to the high effort put into the task by the other (Kruger et al., 2004). Furthermore, goal contagion is a highly social process that requires human interaction (Palomares, 2013). Human interaction happens more often between in-group members (e.g., entrepreneur and employee of the same firm) than among out-group members leading to goals being more contagious inside the same group (Loersch et al., 2008).

The theory of social motivation supports the theory of goal contagion inside social groups. It states that individuals in social situations behave reciprocally in collective tasks meaning that if a person does something for someone else, the other feels obligated to return the favor (Breugst et al., 2020; Geen, 1991). Following social motivation theory, first, the employee will engage in social comparison processes with the entrepreneur and thus try to match the invested effort. Second, he will try to comply with the standards set by the entrepreneur and view his behavior as a benchmark for his own.

Emotional contagion

Let us take the example from above again and imagine the hard-working entrepreneur together with his employee in one room. While perceiving the entrepreneur, the employee will not only notice his behavior, but he will also notice the entrepreneur's emotions as he is putting effort into his work. When the entrepreneur is doing well on his tasks and enjoys them, he is likely to express his positive emotions, for example, by smiling or by communicating the successful progress. The employee can interpret these emotions as the entrepreneur's passion for his entrepreneurial tasks. He will deduce that the entrepreneur must be passionate as he spends hours of effort on his work. Due to their close work, the employee will again infer that he is in the same boat as the entrepreneur and will experience similar feelings for his tasks. As the employee can perceive the emotional reaction of the entrepreneur at the workplace, the theory of emotional contagion states that the feeling of passion can transfer from the entrepreneur to the employee, similar to goal contagion. Daily life shows that if you are engaging with someone who expresses his positive emotions, laughs, cheers, and is in a good mood, it is likely that his feeling will catch you.

The literature converges on emotional contagion as a flow of emotions from one person that others can catch. Scholars agree on this process as "the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person's and, consequently, to converge emotionally." (Hatfield et al., 1993, p. 96). Therefore, the common understanding of emotional contagion is that people can "infect" (Elfenbein, 2014, p. 327) other people with their emotions. Hatfield et al. (1993) provide two possible mechanisms for emotional contagion. First, mimicry takes place. The example of yawning can best explain mimicry. When we see another person in the room yawning, we likely start yawning just because we perceive the other doing so. Mimicry describes the human tendency to mimic another facial expression, gestures, and vocal utterances, which starts at an early stage when a newborn mimics its mother. Observing children mimicking their parents is also an early test for autism, as this diagnosis often follows from misinterpreting emotions (Helt et al., 2020). Mimicry can also happen in a very subtle, unconscious, and rapid (less than 21 milliseconds) way, where it is not observable with the human eye but can be tracked with technology as it takes place. Therefore, people can automatically mimic other people's emotional characteristics. Second, feedback occurs. Feedback means that during mimicry, the central nervous system sends signals to the brain, letting people make inferences about their emotions based on their behavioral expressions. Expressions and emotions are linked as when people, for example, express emotions such as being happy or sad with their faces, they are likely to feel the expressed emotion. Therefore, "emotions are shaped by feedback from posture and movement" (Hatfield et al., 1993, p. 98). In our context, an example of emotional contagion is when the employee perceives his supervisor as passionate by smiling. He will automatically mimic the smile leading to the feeling of passion via feedback, which leads to him feeling passionate.

Emotional contagion can also occur consciously through social comparison (Barsade, 2002; Elfenbein, 2014). Social comparison describes the process of comparing own emotions with the perceived emotional states of others in one's environment and responding according to what one finds appropriate. An employee working with the entrepreneur on entrepreneurial tasks and perceiving his supervisor experiencing passion while working is likely to compare his emotional state to his supervisor and experiencing passion himself.

Scholars found emotions more contagious if people are in a close or similar situation with one another (Platow et al., 2005; Sullins, 1991). The employee and the entrepreneur work for the same firm and sometimes on the same tasks. They typically share many values and interests. Employees will perceive themselves in the same group as the founder (Breugst et al., 2012). Furthermore, the occurrence of emotional contagion is more likely, if the expressed emotion of the sender is more intense or energetic (Barsade, 2002). For example, suppose the entrepreneur is highly extroverted and communicates his emotions loudly or shows many gestures and facial expressions. In that case, the employee is more likely to experience emotional contagion because he will focus more attention on the high-energy expressions of his supervisor.

It is important to mention that not all emotions are trivial to interpret for a perceiver or transfer equally. For example, in concordant affective transfer, when a person is happy and expresses that feeling through laughing, the perceiver might also experience this positive feeling (Epstude & Mussweiler, 2009). On the other side, malicious joy describes a situation where discordant reactions are displayed. An example could be that a person laughs about another person falling on the ground. So the experience of pain or suffering made the perceiver feel happy, which is a discordant affective transfer as there was no transfer of similar affect but a transfer of negative to positive affect (Heider, 1958). In the entrepreneurial context, the employee could experience malicious joy when the entrepreneur is not performing well at an entrepreneurial task, for example.

2.3.2. Self-Regulation: Control, Goal-Setting, and Social Cognitive Theory

I use theories on self-regulation to explain the intrapersonal path starting from effort that leads to passion. This framework allows us to think of emotions as an outcome of behavior and not as an input factor. Using this group of theory explains that people become passionate about an activity because they put effort into it. The rationale is that people set goals and try to achieve them by completing specific tasks. Coming closer to the desired goal leads to progress. Progress toward a goal lets people experience positive emotions. These positive emotions are typical of passion. Therefore, people become passionate because they try to reach their goals while putting effort into work tasks.

Scholars used control theory to explain self-regulating systems in academic fields like engineering, applied mathematics, economics, or medicine. Due to the breadth of the appliance, it developed into a general theory for self-regulation systems, and finally, Carver and Scheier (1982) used control theory to describe human psychology. They argued that people act in a negative feedback loop. Negative means that people try to diminish a disequilibrium between the current and desired state. People start this loop by perceiving the current and initial conditions in the present. Then, they compare the current state with the desired goal state. If the comparison results in inequality, they initiate a behavior to close the gap between the as-is and the to-be state. If people put more effort into their tasks, they will likely close the gap between the current state and the desired goal state with a higher rate of progress. Carver (2006) stated that reaching a goal at a higher rate will develop the experience of positive affect. These positive feelings mean that "you're doing better at something than you need to (or expect to)" (Carver, 2006, p. 106), and as explained earlier, positive feelings are typical for passion (Chen et al., 2009).

The theory of goal-setting supports the presented idea. Locke and Latham (2002) believe that goals provide the following mechanisms: First, setting goals enables a clear direction for attention and effort. People tend to put more effort into what is relevant for goal achievement than goal-irrelevant activities. Second, goals energize people. The higher the goal, the more effort people put into their tasks. Third, goals increase the persistence of people working on a task meaning that more challenging goals lead to a more extended period in which people can direct their effort. Finally, goals are connected indirectly to the knowledge created that is relevant to the task. Furthermore, setting goals creates a discrepancy between a current state and a reference value desired to attain (Locke & Latham, 2006), which goes hand in hand with control theory. Besides, Locke and Latham (2006) found out that achieving goals is connected with the experience of positive emotions toward the task, as goals determine how satisfied people are with their work.

Similarly, social cognitive theory supports the described idea and links to control and goal-setting theory. The rationale of social cognitive theory is that the more effort people put into their tasks, the more progress and success they will make, which stimulates their experience of positive feelings toward their tasks (Bandura, 2001; Bandura et al., 1999). Stimulation occurs through a higher level of self-efficacy, which leads people to anticipate success and progress on their work tasks (Bandura, 1988). Social cognitive theory suggests and substantiates the causal chain from effort to passion.

A noteworthy mention is that not only the fulfillment of the final goal is associated with progress and thus experiencing positive emotions but also the completion of subtasks (Gielnik et al., 2015). Weick (1984) noted those complete sub-tasks as "small wins" that help reach the ultimate goal. The accomplishment of sub-tasks indicates that there is substantial progress towards the goal, which leads to a positive effect on the emotions of the people involved. Concerning my study, an ultimate goal could be the overall success of the new venture in terms of financial performance. A small win could be a positive call with a potential investor securing a follow-up interview.

In our context, the startup employee will elaborate with the entrepreneur on specific entrepreneurial goals and tasks to work on them. This collaboration will start the discrepancy-creating process as after they identify the tasks, the employee has a goal state to achieve. To close the discrepancy between the as-is state and the to-be state, which is completing the tasks, he must mobilize effort. As entrepreneurial goals are higher-level goals, the employee will increase his effort. He will progress and succeed during his work, experiencing positive feelings typical of passion and becoming passionate about entrepreneurial tasks. This process is valid for accomplishing the ultimate goal but for every successful subtask, including small wins.

2.3.3. Combined theoretical framework

Reviewing the literature leads to the necessity to combine theories of contagion and self-regulation to explain the relationship between perceived supervisor effort and entrepreneurial passion. Figure 1 summarizes the theoretical framework provided in this paper. This framework combines different research fields to capture how the mere perception of entrepreneurial effort affects the employees' entrepreneurial passion response.

The starting point of this system is the displayed supervisor effort. As he works on entrepreneurial tasks, he displays his passion, which others can perceive. This process resembles the first path of this theoretical framework: the interpersonal transfer of entrepreneurial passion (Path I). With the theory of emotional contagion, scholars explain that employees can become passionate about entrepreneurial tasks if they perceive the passion displayed by their supervisors. The second interpersonal path of this framework describes the bridge between the entrepreneur's and employee's efforts (Path II). Goal contagion theory explains that employees can infer the goals of their supervisors and pursue them by themselves. Perceived effort acts here as a catalyst that enhances the likelihood of this phenomenon, but there is also evidence that effort is contagious itself and can transfer directly. The third path of this framework is intrapersonal, as it describes how entrepreneurial effort can trigger entrepreneurial passion (Path III). It is essential to mention that this mechanism works for both the employee and the entrepreneur. Path III can therefore initiate and explain the connection of entrepreneurial effort to Path I as entrepreneurs become passionate while working effortful on their tasks and thus display higher levels of passion. The same is valid for employees. They can develop an entrepreneurial passion when they put effort into entrepreneurial tasks. The theory of self-regulation illustrates this phenomenon.

To conclude, this theoretical framework provides the interplay of three different research streams to deliver possible

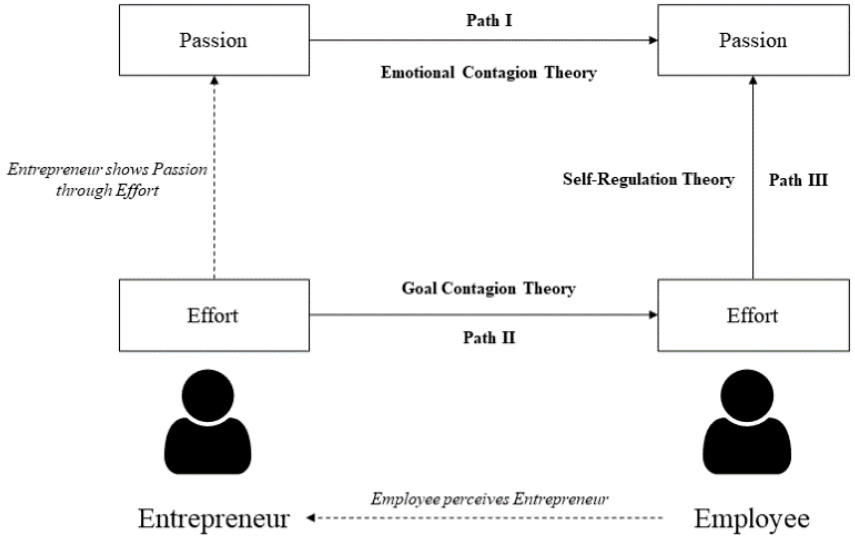


Figure 1: Combined theoretical framework of contagion and self-regulation

explanations of how the employees' passion response is affected by the perception of their supervisor's effortful behavior. The current study aims to extend this framework. Past academic contributions on this topic have mainly focused on experimental or quantitative approaches to find evidence for causality for their claimed hypotheses or to measure changes in the mentioned variables. With an explorative approach, this study intends to close a research gap and put the entrepreneurial employee into the focus of analysis to find support for the provided framework, discover alternative paths or further factors, and investigate boundary conditions for the presented relationships.

3. Methods and data

3.1. Research design

I used a qualitative approach to answer my research question. Qualitative research is a set of field-based methodologies with the participant at the center. Qualitative research aims to generate or develop a theory rooted in first-hand observations. It appears in natural settings, is flexible, reflexive, unstandardized, and the data emerges from the participant's perspective (Lee et al., 1999). Lee et al. (1999) describe qualitative research as a method of data reduction that concurrently improves the significance of the data. According to them and other academics (e.g., Pratt 2009), besides solely describing and documenting what is occurring, qualitative research can explain the how and why to understand the big picture of a process in reality. Therefore, this approach is effective in environments with at least limited theory or knowledge of how a process works, like in this current research. As previously described, there is already existing theory explaining independent mechanisms in the researched process. However, there are still gaps in how and under which circumstances these components work together. Furthermore, there are no formal propositions present. Therefore, this paper aims to further develop a theory on this stream of research in entrepreneurship resulting in specific research propositions (Edmondson & McManus, 2007; Lee et al., 1999).

I utilized a multiple case study approach that helps to gain new insights into a less-investigated research subject to add novel insights and observations about the academic topic to the current state of research (Eisenhardt, 1989). Since there is limited theory and evidence on how perceived supervisor effort influences the employees' entrepreneurial passion, a case study approach is well-suited as a methodology. Furthermore, the advantage of using multiple cases compared to a single case study for developing theory is that a multiple case study approach establishes a more robust theoretical foundation with higher richness and accuracy of theory (Yin, 2018). The theory that has emerged from several case studies follows a replication logic and is therefore "more grounded, more accurate, and more generalizable" (Eisenhardt & Graebner, 2007, p. 27), which is supported by other scholars too (e.g., Gehman et al. 2018). A multiple case study approach can also respond to questions that start with "how" or "why" (Edmondson & McManus, 2007; Gehman et al., 2018), and it is also appropriate when there is a process-related research question and when variance emerges throughout the data (Langley & Abdallah, 2015). In order to investigate similarities and differences across my cases, I performed a series of case studies to obtain a variety of data.

To gather data, I conducted semi-structured interviews. According to Edwards and Holland (2013), semi-structured interviews offer flexibility for both parties. They resemble a typical tool in qualitative research to get the most out of interviewees. Compared to more conventional experimental or survey approaches, interviews better match the study's theoretical question and analytical context (Lee et al., 1999). Semi-structured interviews allow for more spontaneous inquiries and a more narrative discourse than structured in-

terviews, which are composed of questionnaires with a set order of questions that must be asked one after another (Edwards & Holland, 2013). Additionally, fluid dialogues that include interactions are appropriate for semi-structured interviews (Mason, 2017). The semi-structured interviews enabled the employees to freely share their perceptions and experiences while also allowing me to compare the results with those from later interviews and form conclusions. These factors led me to conclude that semi-structured interviews were the best strategy for collecting rich data for the research topic.

I employed a theoretical sampling strategy to concentrate on developing existing theory (Miles & Huberman, 1994). This strategy is ideal since I want to gain the most theoretical insights possible for my study topic by selecting the appropriate population (Eisenhardt & Graebner, 2007). Eisenhardt and Graebner (2007, p. 27) stated that "theoretical sampling simply means that cases are selected because they are particularly suitable for illuminating and extending relationships and logic among constructs," which is the goal of this study. Additionally, a theoretical sampling strategy is helpful in the search for data with high information content (Patton, 2009).

Even though most academics connect qualitative research to inductive approaches and quantitative research is connected to deductive approaches, other researchers claim that qualitative research can be either inductive or deductive or a combination of both (Pratt, 2009). The presented knowledge from the literature concerning this research question does not propose an already investigated tested mechanism of how exactly perceived supervisor effort affects the employees' entrepreneurial passion. Instead, it combines different research streams and forms a temporal theory. Therefore, the study has to test the proposed framework with the data gained. According to Bitektine (2008), case studies are applicable to test for theory as they can be seen as similar to single experiment tests. As the combined temporal theory on contagion and self-regulation derived from the literature does not intend to fully describe the effects of perceived effort on the employees' passion for entrepreneurship, I seek to extend the current theory for this specific body of research. However, the data must confirm whether the assumed relationships and constructs can occur in the described setting. As case studies help build, develop and test theories (Eisenhardt, 2020), they are valuable for this study because some of the applied theories must be confirmed before developing new relationships.

As I worked inside a theoretical framework and wanted to develop this framework further, I started initially with a top-down coding approach. I deduced some codes from the research question and theory and worked inside these boundaries as openly as possible. At the same time, I kept an eye on leaving the boundaries when new knowledge appeared outside the current framework. Then, I mainly used an inductive approach recognized in qualitative research (Strauss & Corbin, 1998). I could look for differences and similarities in the data gathered by using this coding strategy which gave me a clear direction in the data analysis process.

3.2. Sample and data collection

I took numerous steps to guarantee proper sampling and adequate variation between the cases and the selected interview partners. I used theoretical sampling to interview participants currently or previously employed at a startup. Besides, they should have worked closely with the founders to gain a proper perspective for perceiving entrepreneurial effort. With this careful case selection, I increase the probability of observing the "focal phenomenon, mitigate alternative explanations, and enhance generalizability" (Eisenhardt, 2021, p. 149).

Additionally, specifying the population constrains the extraneous variation for the phenomenon and sharpens external validity. By working closely with the founders, I mean that I looked for employees who were either at a high hierarchical level inside the firm or joined the firm instead in the early stage (less than 100 employees and five years old). The latter is because when startups have few employees, they often tend to have shared workspaces in small offices with the founders, and so they perceive them intensively. Interviewing participants who used to work for startups but are not employed in small ventures anymore allowed me to draw deeper inferences about the outcome of the process rather than the process itself, which is also essential to this study as this leads to "better grounding and external validity" (Bingham & Eisenhardt, 2011, p. 1440). Also, combining retrospective and real-time cases mitigates bias as they efficiently enrich the cases by quantity and depth (Eisenhardt & Graebner, 2007). I chose to keep my sampling strategy open to various sources to incorporate different employees. Eisenhardt and Graebner (2007, p. 28) explained the importance of "using numerous and highly knowledgeable informants who view the focal phenomena from diverse perspectives. To limit bias, these informants can include organizational actors from different hierarchical levels, functional areas, groups, and geographies [...]". With this sample selection strategy, I could control external variation while paying attention to the instructive variation of interest.

My sample consists of eight different startup employees who gained experience mostly in firms founded in the south of Germany. These eight interview partners serve as foundations for my cases. They are the primary data sources, consistent with using the multiple case study approach (e.g., Breugst et al. 2015). With eight cases, I am in line with the typical number of cases used for multiple case study research design (Eisenhardt, 1989, 2021). I approached the sample through my network, references from people in my social network, and suggestions from the interview partners at the end of the conversation. I had been acquainted with five out of eight interview partners before the interview already. Two participants work for the same startup, which allowed me to compare their perceptions of their supervisor's effort. Except for one firm that produces hardware for the medical industry, all other firms develop software for different industries, for example, gastronomy, construction, and logistics. Seven out of eight participants were native German speakers, so I translated the interview questions for them and conducted these

interviews in German. Table 1 shows an overview summary of the sample cases.

After approaching the interview participants initially, primarily via phone or messaging via LinkedIn, I got their agreement to participate as interview partners in my study. Then, I made an appointment for the interview within the next 14 days while letting my participants choose the exact time and date. I did this to ensure that they had enough time for the interview and no tight schedule or other appointments that applied pressure to the interview timeframe. Besides, I wanted to ensure that the interviewee's responses were not affected by these factors due to this flexibility in time and location through the video call (Lee et al., 1999). Throughout the initial approach, I asked their permission to video-record the interview and ensured confidentiality and anonymity for the presented data in the final paper. This promise enhances the honesty and integrity of the participants (Huber & Power, 1985). Then, I sent them an email invitation for their calendars with the corresponding link to enter my virtual video room. I blocked one hour for the interview to account for any delays and have enough flexibility to arrange the conversation within this timeframe. The only information I exposed about the study to the participants was that in my research: "I investigate relationships between employees and entrepreneurs within new ventures.". Therefore, they were not biased or primed by exposing too many details about my research or knowing the research question.

After starting the video call, I reminded them about the video recording to ensure transparency and reminded them again that all personal data would be anonymized. Additionally, I promised them to stop the interview whenever they did not want to continue. I encouraged my interview partners to speak out as openly as possible and told them there were no right or wrong answers as I was interested in their experiences and opinions. In the end, I asked whether I could contact the interview partners afterward in case of any questions. Additionally, I asked if they could suggest someone else who would fit into my study. This approach led me to five recommended interview partners, two of whom further agreed and participated.

As previously stated, the research builds on semi-structured interviews that I performed using the technique of Eisenhardt (1989). I based my methodological approach on qualitative studies addressing entrepreneurship using the same scientific methodology (Breugst et al., 2015). I was able to extract the most information possible from the cases thanks to the ability to modify the interview questions based on the progress of the interviews. Sometimes I got quick responses to the questions I had meant to ask. By asking more detailed questions or focusing on particular episodes in the participants' stories, the semi-structured setting of the guideline allowed me to go deeper into these cases. Recall bias from the interviewees was a common concern prior to conducting interviews. I tried asking interviewees about specific incidents or memories rather than their general perceptions to prevent this bias (Podsakoff & Organ, 1986). Event tracking was a crucial aspect of the data gathering, in addition to avoiding bias, because explicit statements about events helped me find patterns in the data.

In order to prevent bias in the composition of the questions, I created the interview guidelines for the semi-structured interviews before obtaining or considering concrete employees as participants. According to how the conversation developed, I only utilized pertinent questions that were largely interchangeable, even within sections (Edwards & Holland, 2013). I adjusted the number of questions for relevance to the interview partner. Generally, the interview guide contained 34 questions, with 11 mandatory and 23 optional. I structured the guideline into five main sections, with a 3-level hierarchy starting each section with opening questions I asked on the first level. The second level was about to specify the opening question further. The third level was mainly on a yes-or-no level to get a more profound and precise understanding of the interviewee's response. I asked the first-level questions in every interview. Level 2 and 3 questions were possible follow-up questions to guide the conversation in a more directive way and were more or less optional to the conversation.

In the first section, I asked for a brief and general introduction of the employee and the startup they currently or formerly worked in. This opening question helped each participant familiarize himself with the interview environment and let me learn more about the participants' experiences. It also gave me an overview and a preliminary understanding of how the employee sees himself and his business. Additionally, this assisted me in my data analysis since I was getting to know the different stories. The second section of questions concerned the participants' entrepreneurial passion and effort. I asked them to capture the employee's general level of entrepreneurial passion and effort, as these are core concepts of this study. The third section was about the relationship between the entrepreneur and the employee to better understand their interactions and communication. The fourth and fifth sections contained questions about the perceived passion and effort of the entrepreneur and the possible effects of perceived effort on the employee's entrepreneurial passion, as this is the core question of this study.

The final interview guideline helped me with the following: First, I verified that every interview partner fulfills the necessary criteria to participate in this study, for example, if they have or had enough interaction with the founders to have an accurate perception. Second, I tested some essential theoretical assumptions, for example, if employees perceive to be in the same boat as their founders. Third, I gained new knowledge by asking about concrete effects and possible explanations for the phenomenon. To provide transparency, I attached the complete interview guide in Appendix A1 and the German translation in Appendix A2. I explained every question and referenced literature when applicable to enhance internal validity.

The interviews were conducted via video calls using the software Zoom, which allowed me, together with the approval of the participant, to record the video sessions. Therefore, I gained richer data due to not only having the audio

Table 1: Summary of sample cases

Employee name	Employment status in small venture	Role	Period of employment	Industry	Founding Experience	Known from personal network
JG	Prior employment	Legal Consultant	2017-2018	Data Security	None	Yes
GD	Current employment	Software Developer	2018-today	Healthcare-Hardware	Prior venture founded	No
RB	Current employment	Venture Developer	2022-today	Construction-Software	Prior venture founded twice	Yes
MS	Current employment	Software Developer	2021-today	Logistics-Software	Prior venture founded	No
CH	Prior and current employment	Customer Success Representative	2018-2019	Identity Management	Subsequent venture founded	Yes
JK	Prior employment	Product Manager	2017-2022	Gastronomy	Subsequent venture founded	Yes
LL	Current employment	Chief of Staff	2022-today	Construction-Software	Prior venture founded	No
AW	Current employment	Business Development Representative	2021-today	HR-Management	None	Yes

files but the corresponding visual data. Recording helped me a lot since I could omit taking notes frequently while doing the interview and focus on the conversation, which made the interviews more authentic and enabled me to establish more trust with my interview partners. With the video recordings, I accounted for peoples' gestures, facial expressions, and emotions. Consequently, I better understood the conversation afterward and enriched the spoken word by further interpreting and perceiving visual data. Having video data of the conversation is, therefore, superior to recorded audio only, which increases the reliability of this study.

Before starting with the regular interviews, I conducted a pilot interview with a close colleague to check if the technical setup was working correctly, without connection or audio issues and if the exported files were processable for me afterward. Additionally, I tested if the questions were understandable without explaining them deeper. Finally, I could use the pilot interview to get used to the setting, account for good light and audio conditions, and have more confidence in asking the questions after testing them once. The regular interviews benefited much from the pilot interview since no significant troubles or disturbances occurred.

Regarding the duration of the interviews, I set a goal of approximately 30-45 minutes because I believed this was required to glean insightful information and to follow academics' recommendations to make prior decisions to the wished interview duration (e.g., Lee et al. 1999). The recordings contain 273 minutes of video data. Transcribing the interviews resulted in 71 single-spaced pages. For transparency reasons, I included the transcripts of all interviews in Appendix A4. Within the next 24 hours of doing the interviews, I transcribed them. As a result, there should be a better assessment of the cases and a deeper engagement with little loss of significant interview-related memories. On average, the interviews took 34 minutes.

I stopped seeking new interview partners and collecting further data as I began to receive the same information that only added little benefit to the study, which researchers call the point of reaching theoretical saturation (Lee et al., 1999). Also, as I iterated between data analysis and collection, I anticipated no new codes for the final coding scheme after conducting the eighth interview, suggesting that theoretical saturation had occurred (Aguinis & Solarino, 2019).

In addition to gathering information from interviews, I also looked for secondary data to back up the collection and corroborate the triangulation method (Jick, 1979). After finding agreement across the interview data, triangulation in qualitative studies can show "agreement among different sources or types of data" (Lee et al., 1999, p. 179) and further validate the gained results. Edmondson and McManus (2007, p. 1157) describe the strategy of triangulation as a "process by which the same phenomenon is assessed with different methods to determine whether convergence across methods exists.".

Therefore, I gathered information from the startup websites and CrunchBase, watched videos, and read articles

about their teams and products on YouTube, Instagram, and LinkedIn. Besides, I looked at the supervisors' and interviewees' educational and professional backgrounds on LinkedIn and other material I found about them, like their websites or blogs, publications, articles in certain magazines, or interviews they conducted to enhance my understanding of their personalities. Additionally, the participants showed me presentation decks of their respective companies, which I was allowed to take notes on but not to use afterward in my data analysis as a separate file. Finally, I looked on the internet for further interviews with employees of the respective firms for deeper insights and validating statements. To collect these data, I utilized Google as a search engine and applied a keyword strategy using operators such as "and" and "or" to account for aspects and synonyms. Then, I created a matrix for defining and tracking search terms to combine different wordings and languages to create specific search queries. To form a search term, for example, I took the name of one of the ventures and added the words interview and employee with the connector "and" to look for employee interviews of this firm. I used a webpage-to-pdf converter (Print Friendly & PDF) or the print function of the Google Chrome web browser to scrape the information from the different sources and make it usable for data analysis software later in the process. For video or audio material, I used a software named Trint which delivers artificial intelligence-based transcripts with a sufficient accuracy rate to save time in the data collection process and make it usable for further analysis.

This kind of data collection technique is commonly accepted and used by scholars (e.g., Gehman et al. 2018). Secondary data helped verify the interviewees' self-reported claims and the characteristics of their entrepreneurs and increased the overall confidence in the accuracy of the findings. It supported me in understanding the participants' working environment as they sometimes needed help to give certain information on a specific topic because they could not remember or did not mention it during the interview, although I asked them. Table 2 summarizes which data source I used for every case.

3.3. Data analysis

Although I knew I wanted to identify patterns relating to how perceived effort affected employees' entrepreneurial passion while coding, I tried to approach the data with an open mind and no preconceived notions (Suddaby, 2006). I could therefore find notions that were related to the research issue. When reading through the transcripts and secondary data for the first time, I decided against immediately starting the analyzing process because I wanted to become more familiar with the data and get a better sense of the big picture. It was necessary because one challenge was avoiding bias while processing the collected data (Eisenhardt, 1989).

In order to make sure I did not miss anything, I started coding as soon as I finished reading through the data for the second time. To confirm the codes, I coded each unit of data twice. I mixed bottom-up with top-down coding as I do not generate new theory but work in existing theoretical frames and want to allow the narrative to emerge inside this frame from the raw data. Therefore, I anticipated some codes derived from the research question and the theoretical framework I work in that I might use later for building my model and started with the most core concepts of my research: perceived entrepreneurial effort, perceived entrepreneurial passion, entrepreneurial effort, and entrepreneurial passion of the employee. Then, I organized the data inside this categorization with an open coding strategy (Corbin & Strauss, 1990). I coded each statement I believed to be helpful to deepen the insights and subsequently be grouped with other statements and secondary data. By identifying and grouping text units belonging to the same concepts, I allocated codes to statements.

Labeling statements produced numerous first-order codes. The first-order codes' complex narratives are the foundation for a more theoretical and analytical view of the data than just a descriptive one. To organize the first-order codes which emerged from open coding, I followed an axial coding approach (Strauss & Corbin, 1998) to find differences and similarities between the categorized data parts. I processed this by connecting the first-order codes and grouping them into categories that produce a logical whole. It reduced the number of codes and resulted in a better overview of the information pieces. In this step, I sought to aggregate codes into higher-order concepts to pave the way to the corresponding literature and theory. It resulted in the creation of second-order codes. An investigation of this kind enables the discovery of potential underlying dimensions or patterns in the data. Next, I used selective coding to the emergent patterns in the data to extract the theoretically explanatory dimensions (Strauss & Corbin, 1998) and build the bridge back to my initial coding scheme considerations. The final phase of my analysis started with abstracting themes into higher-order theoretical dimensions. I iterated back and forth between my interpretations and the data to ensure the former held.

I iteratively worked on the data coding using MAXQDA (Version: Plus 2022, Release 22.3.0) as my comprehension of the research issue grew and deepened (Strauss & Corbin, 1998). As a result, whenever I developed new insights into the material, I had to re-check the coding scheme. It involved renaming first-order codes and re-clustering my codes into subcategories. This process follows the replication logic for the multiple case approach, as every single case can be observed independently and not as an additional data point (Eisenhardt, 2021). Therefore, I tested every case for the emergent theory's occurrence, which helped me gain familiarity with data and preliminary theory development. It also helped to identify patterns across the cases. Observing every single case on its own encouraged me to look past my first impression and consider the evidence from various angles, to form relationships and investigate the underlying mechanism. I updated the codes regularly during the analysis following this procedure. It means I added new codes while dropping other codes in the process. During data analysis, I

Table 2: Overview of used data types for the sample cases

Case	Interview Data	CV	CV Founders	Website Information	Other Employee Interviews	Articles	Founder Interviews	Other information
JG	8 pages	Yes	3 CVs	37 pages	28 pages	-	-	-
GD	8 pages	Yes	4 CVs	35 pages	-	4 pages	-	6 pages
RB	10 pages	Yes	3 CVs	8 pages	12 pages	6 pages	-	64 pages
MS	10 pages	Yes	3 CVs	13 pages	-	-	13 pages	18 pages
CH	8 pages	Yes	4 CVs	17 pages	-	-	16 pages	26 pages
JK	10 pages	Yes	3 CVs	9 pages	-	11 pages	-	1 page
LL	10 pages	Yes	3 CVs	8 pages	12 pages	-	-	4 pages
AW	7 pages	Yes	2 CVs	47 pages	-	7 pages	3 pages	-

regularly iterated between the emergent theory and the data to compare existing knowledge with new information to close the research gap, as recommended by Eisenhardt (2021).

Overall, the coding and analysis process was done iteratively and repeatedly before ending up with a final scheme and analysis (Miles & Huberman, 1994). This strategy helped me gain new insights from the data I might have missed if I had only gone through it once. Besides, I asked a colleague to check and agree on the meaning of the codes and the correspondent data. Double-checking enhanced the accuracy of my findings. Together we re-evaluated the data, committed to discussions, and converged interpretations if there were discrepancies over specific codes. As the last step, I compared all data to the final coding scheme at the end of the analysis process. Figure 2 shows an excerpt of the data structure scheme. The complete list of codes, with explanations and examples for every code, and the code structure, is shown in Appendix A3.

4. Findings

To assure the promised anonymity and confidentiality for the interviewed people and their respective firms and founders, I replaced their names with acronyms of randomized names, for example, JG derived from John Grey. When I write about the company that JG works for, I refer to it as JG_Company, adding an underscore and the word company to the respective name. The same holds for the founder of the respective firm, who will be referred to as JG_Founder if there is only one founder, and if there are multiple ones, I will add a number suffix to the name, for example, JG_Founder_1 representing one founder and JG_Founder_2 representing an additional founder. The numbers resemble an order of proximity to the founder, meaning that JG worked closer with JG_Founder_1 than with JG_Founder_2. I used this naming convention for every data related to the correspondent interview partner.

4.1. Case descriptions and within-case analyses

4.1.1. The case of JG

JG is 28 years old and studied computer science and business administration in his bachelor's and master's programs at a large university. During his studies, he wanted to explore the everyday life of a startup and joined a young venture for one year overall as his first job. The firm operates in data security. They consult within these topics as and offer small software solutions to provide the regulatory standards that small and medium-sized companies need. The firm was founded by another firm that hired three experienced CEOs to build and grow the startup. As JG joined the company in its first year after its foundation in 2017, his tasks were broad. His primary focus was working in sales and generating as many leads as possible to acquire new potential customers for the firm securing a deal with them. Because of his IT background, he could support requirements analysis for software selection, processing incoming trouble tickets, and even programming. Overall, he acted additionally as an assistant to the CEOs as he supported them with their daily work and even had to search for another office as the firm started to grow. JG cooperated mostly with JG_Founder_1. She brought in much working experience as she had previously worked as a business analyst in consulting and as a head of corporate sales for another startup in the food industry. In the end, JG left the company to explore another challenge at an IT consultancy, where he still works nowadays and is enjoying his job. After working for the startup, he never got in touch with other small ventures.

Before working for this company, JG had no touching points with startups or entrepreneurship. He mentioned that generally, he would like the idea of founding his own company one day if he finds an appropriate IT-related product and could imagine taking the necessary tradeoffs like not earning much in the beginning. All in all, he started with a low passion for entrepreneurship in comparison to the other participants.

JG perceived his founders to put high effort into their tasks and the venture. He reported that his founders always were willing to go the extra mile "without exception" and that he has "never seen them going home or coming to work ever" due to the long working hours he perceived. Besides, he received emails and task descriptions on the weekends and estimates for his supervisors additional "12 hours or 8 hours normal working time on Saturdays and Sundays". He perceived high entrepreneurial passion from his founders. He

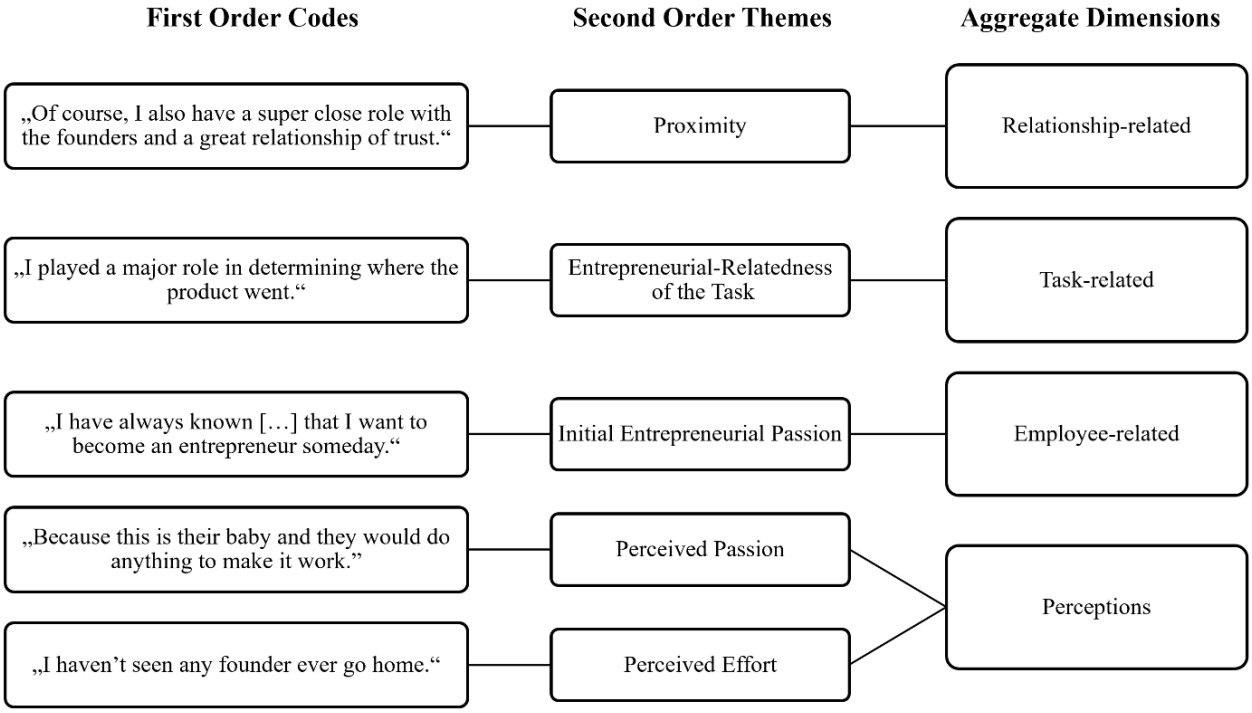


Figure 2: Excerpt of data structure

described JG_Founder_1 as “very passionate and very extroverted” who always had to “drop and leave everything else aside” when she had an idea and pushed for immediate implementation as she “really fell in love with her ideas and pushed them towards the end.”.

JG claimed that this behavior influenced him as he “therefore stayed often longer in the office” to collaborate with JG_Founder_1 on her ideas. He reasoned that he increased his effort because of his passion for the tasks. He also mentioned that progress made his supervisor happier, and “she seemed to be very interested in this progress. That was actually the passion that made [him] stay longer because [he] knew that it was very important to her and that she was extremely happy when there was progress.”. Overall, JG increased his level of effort while working for JG_Company as he “was almost always willing, if [he] had no other obligations […] to go the extra mile and also stay late at the office.”. However, he did not enjoy this amount of effort after a few months because he either “wanted to do other stuff,” and “had other interests than sitting in the office with the founders after 11 pm” or disliked the expectancy of “continuously staying until 11 pm the next day too”. He realized that this is not the working schedule that fits in his life and therefore chose to join a firm where he had a stable 40-hours-week. He described that this experience “brought [him] further away from founding a venture” which fits the fact that he never showed interest to the entrepreneurial process afterward. To sum up, JG perceived much entrepreneurial effort and passion, which made him increase the effort he put into his tasks while working with the founders. However, this amount of effort decreased his passion for entrepreneurship.

4.1.2. The case of GD

GD is 29 years old and was referred to this study by JG from his network. GD earned a degree from a Turkish university in computer engineering. Her long-term goals, which she also mentions on her social media accounts, are to start a business, become a strong tech woman, and create cutting-edge technology. After working as a software developer for a Turkish company, she moved to Germany four years ago to start working for the GD_Company, a startup operating in the medical sector. There, she is still working as a software developer. She joined the firm as one of the first ten employees. GD saw the startup grow to 450 employees, but a Swedish company acquired the company during this period. The initial founders dropped out after the acquisition and got replaced by two managers. GD_Founder_1 and GD_Founder_2 founded the firm in the first place with more than 30 years of combined working experience in relevant fields and sectors. In comparison, GD_Founder_2 brought expertise in business for the medical industry, and GD_Founder_1, as CTO, was the primary contact person for GD discussing relevant technical topics with previous experience in relevant research fields. GD_Founder_3 and GD_Founder_4, who bring equivalent experience to the firm, replaced the founding team.

GD had her first touching point with entrepreneurship before working for GD_Company. She participated in a programming competition where she won. By winning the competition, she got a sponsorship to build her own company, which she tried to start. According to her, the later business failed because the sponsorship was insufficient, and the government refused to support her further, making her look out for international opportunities. She discovered an opportu-

nity for support in Germany, but at the same time, she got a job offer from GD_Company, which she ultimately took. Nevertheless, GD started her job with a high initial passion for entrepreneurship as "it was [her] main idea to found a startup.". She reported that when asked during her job interview where she sees herself in 5 years, GD said she wanted to have her own business then.

GD enjoyed working with the initial founders before they left the firm. From GD_Founder_1 and GD_Founder_2, she perceived a high level of effort. She had never seen them leaving the office before herself and perceived them to invest extra effort beyond what was immediately required. If there was a problem at work, "they had to solve it today.". After the initial founders left the firm and the new managers took over, she perceived a decline in effort: "so the founder team, I would say they were more invested as the current ones.". She described the initial founders, GD_Founder_1 and GD_Founder_2, as "the creative minds" who brought the ideas in and worked intensively to improve the product. However, after the change, she saw GD_Founder_3 and GD_Founder_4 "working the same as us - 40 hours" per week, not being committed to improving the product as they "don't bring the ideas" and leaving the office regularly before herself. Additionally, she perceived low passion from the new supervisors. According to her, they focus on profitability, company benefit, and delivering a minimal valuable product. Instead, the initial founders were able to create and improve the product and cared more about the customers as they traveled more to customers in person to solve their problems.

Perceiving different levels of effort and passion by the two different managing teams led to different effects on GD. Perceiving the effort of the initial founders increased her level of effort while they worked in "challenging" times and solved problems together. They enhanced her in gaining responsibility for more entrepreneurial tasks when she traveled to customers for specific tasks even though "normally, software developers do not travel." Especially the effort they put into their product affected GD as she said, "[. . .] if your employers are more focused on entrepreneurship or they are spending more effort on the product that they have created, you see that, and you also get more motivation as well.". This effect changed to the opposite after the new managers were in charge, and GD perceived little effort. Now, she is considering reducing her working hours from 40 to 35 and tracking her working time to avoid substantial overtime. She will not increase her effort if she does not benefit personally from it. Concurrently, her entrepreneurial passion decreased as she now feels "that I kind of lost that mindset as well. But with the founders, I was so much searching to do the stuff, and I was so really focused on the idea of having a business, but right now, I lost it. Not fully lost it, but it is definitely decreased. So, there is a huge effect.". She also perceived this effect in her last job when she worked solely on solving daily problems and lost the drive to continue on entrepreneurial tasks. Narrowing down her effort solely on daily tasks was "kind of affecting" for her, "even though you don't realize it.". She feels the same as she had felt in her previous job: "Being a normal employee that's just saving the day as I had in Turkey."

To explain this effect, she used a metaphor related to the contagion effect described in the theory section: "In Turkey, we say if you want to know the person you should ask the friends. You look at the friend to understand the other person. So, if you are working with a colleague who is entrepreneurship-focused, you will be mirroring yourself one day. As soon as you get closer, you communicate, you discuss, so you get some ideas from your friend. So, this is what I feel. So, if you have a friend who is up there, it pulls you up." In addition, she said: "I mean even for the children. If you look at the children, they are the copy of their families. It is the same. You are so close to your family, so you become a copy of them. And the same with friends. You only become friends with people who are like you, or they will affect you in a way that you will look like them at some point. So same for colleagues as well. If they are so motivated and have nice ideas that will bring value to you. At some point, you will be like them. So, it is mirroring or pulling each other to the same level, like this.".

Overall, GD started her journey with a high initial entrepreneurial passion. While working with the initial founders of GD_Company, she perceived a high level of passion and effort, which increased her level of effort and passion. However, after the change in management, she perceived less passion and effort, which decreased her effort and entrepreneurial passion.

4.1.3. The case of RB

RB is a 27-year-old serial entrepreneur who has founded two companies in the past five years. He studied Management & Computer Science in Germany after living in Turkey and Dubai in his teenage years. During a working student job for a cloud service provider for compliance and investor relations, he started his first company, an information platform for cryptocurrencies, with the help of his former CEO after discussing this opportunity with him. RB, therefore, became an intrapreneur. After running out of money during the covid pandemic, he left the firm. Then, he started a mental health company focusing on building a learning and communication platform for people with mental and psychological issues. He reduced to part-time work on this project as he realized that investors were unwilling to invest the money he wanted to launch the product successfully and make the firm grow. He then decided to gain experience in an established young venture to learn from others' experiences.

Since the beginning of 2022, he has worked for a construction-tech startup in northern Germany, which recently secured a Series B investment. The goal of RB_Company is to unite all participants in the construction sector and transform their collaboration through software and customer service. RB started as a venture developer, but he was recently promoted to the interim head of marketing, operating in close contact with the founders. The founding team consists of three people who are still at the company. Together, they bring in more than 25 years of professional experience

in different ventures and gained insights into working for other startups before becoming entrepreneurs.

RB started his position at RB_Company with a high initial entrepreneurial passion as he previously founded two firms. He mentioned that he enjoys working in entrepreneurial environments because of the different impressions he gained during that time. Also, learning new things and being independent "triggered this passion" in the first place, and he believes this passion increases the longer he works in this environment.

RB reported a mixed perception of effort. On the one hand, he recognized the high effort behind the venture by saying: "Even when I look at RB_Company now, for example, I notice how much hard work is behind it, even more than before.", reporting that his founders are "24/7 available" and receiving messages after midnight. On the other hand, he emphasized that his founders try to maintain a balance for their mental health. About RB_Founder_1, he said he "[...] is super athletic and eats a super healthy diet, pays attention to his sleep, and is rather meticulous and disciplined about ensuring he gets enough balance. When asked whether he feels that his founder works too much or too little on his tasks, he answered, "well, that's super different. Right now, I feel like he's taking on a management task more than he's taking on the task. So, he's trying to create more pressure on the people, that it's prioritized properly among them, instead of him doing the task himself.". Compared to the other cases, RB perceived a neutral level of effort. Concerning the entrepreneurial passion he perceived, he was clearer in his statements as he described RB_Company as the founders' "[...] baby for which they would do anything to make it work.". Additionally, he recognizes in every meeting that they show how important the tasks are to them and how much of the "fire" they have. As he also reported feeling the passion spread by the founders, overall, RB perceived a high level of entrepreneurial passion.

As RB founded two ventures before working for RB_Company, it was not surprising that he reported that he constantly works many hours and is willing to go the extra mile "whenever there is a chance to go the extra mile." There was no evidence in the data that his level of effort changed during his time at RB_Company. Neither was it influenced by the perception of the founders' efforts. Concerning the development of his entrepreneurial passion, he reported: "My passion has basically not changed. [...] I simply realized that it was more difficult than I thought. It's more complex than I thought. There are so many problems that you have to manage at the same time." Later during the interview, he once again mentioned that his passion for entrepreneurship stayed the same.

To conclude, RB started his role at RB_Company with a high initial entrepreneurial passion. Although he perceived high entrepreneurial passion and neutral entrepreneurial effort, there is no notice in the data that his level of effort and passion changed while collaborating with the founding team closely.

4.1.4. The case of MS

MS, 25 years old, is a software engineer who holds a master's degree in computer science from a German university. Throughout his studies, he gained working experiences in different company sizes, from small to large corporations, from consulting to production. After graduating, he acted as a Co-Founder and CTO of a software startup in the video stream business for six months. He left the company after the lack of success and personal differences with the other founder. Then, he joined MS_Company, where he also worked during his studies as a working student and intern as a software engineer. There, he builds the foundation for the product together with the CTO. MS_Company is a software-as-a-service company that helps customers connect their systems and platforms. His responsibilities contain the technical onboarding of new customers, maintaining the infrastructure of the software, and working on the integrability of different systems. The founding team consists of three entrepreneurs. MS_Founder_2 and MS_Founder_3 met while working for a large consulting firm after graduating for a few years. MS_Founder_2 started another venture before starting MS_Company, which was acquired just before the new venture creation, and is, therefore, a serial entrepreneur. MS_Founder_1 completes the founding trio and acts as CTO and CPO, whom MS works close with due to his technical background. MS_Founder_1 is highly experienced, with more than 20 years in the industry, and had already former roles as CTO, CPO, and VP in other software firms. MS could imagine trying to found another startup one day.

Because MS founded a startup before joining MS_Company, he can be considered highly passionate about entrepreneurship even before working for his current employer. He reported that he is highly passionate about inventing when talking about entrepreneurial passion. He enjoys creating new products, especially "[...] creating something that other people will use.". According to him, his main drive for his professional career is the "creation of things." He also explained that the startup environment would fit his interest in creating new things best. Hence, he joined MS_Company, where he feels to get exceptional support for his passion.

MS perceives a high level of entrepreneurial effort and passion from his founders. While talking about MS_Founder_1, he mentioned that he could not remember a moment when MS_Founder_1 was not at work. MS perceives MS_Founder_1 to give "1000%" to his company. Besides, he perceives that the boundaries between his founder's work life and private life become blurred since he perceives a high likelihood of getting responses to questions after 11 pm. When I asked MS about how passionate his founders are, he replied: "I think more is almost impossible. That would be my perception.".

While working for MS_Company, both the entrepreneurial passion and effort of MS increased. He described himself as a high performer who puts a lot of passion, time, and energy into his tasks. Also, he is sure that one day he will try the step again to start his own company. He also mentioned that his level of effort is higher than initially expected by his

supervisors, which they also communicated to him. Besides, he estimated that, on average, he works harder and with more effort than other people with similar tasks inside and outside his company. He reported that even in his private environment, people perceived that MS increased his level of effort, for example, when he refused to go out on a Friday evening and chose to work instead. While telling that story, MS revealed notions of contagion as he reported trying to think in the context of the venture and therefore adapted to think entrepreneurially. Furthermore, when asked about the effects of perceiving the behavior of his supervisors, he stated: "I would say rousing and motivating rather - to show that you can do it.", which indicates a contagion effect.

All in all, MS joined MS_Company with a high initial entrepreneurial passion. He perceived highly passionate entrepreneurs who work on their tasks with high effort. Working for his founders increased his level of effort and entrepreneurial passion.

4.1.5. The case of CH

CH is a 28-year-old bachelor's graduate in Management and Computer Science. He gained his first working experience in customer success for a startup called CH_Company, where he worked for one and a half years. He joined the firm during the phase where they secured their Series B round. CH_Company produces software for digital identity recognition and management. CH supported the aftersales process inside the customer success department. The founding team consisted of two co-founders. CH interacted most with CH_Founder_1 and had fewer touching points with CH_Founder_2. Both founders are highly experienced, with more than 40 years of combined working experience. They have already founded and exited a startup in the past successfully together. As a multi-serial entrepreneur founding several firms in their collaboration, CH_Founder_1 has more experience in the field of entrepreneurship, with more than five new ventures created. During his time at CH_Company, he experienced a change in the CEO role. CH_Founder_1 dropped out of the daily operative work and took the role of a chairman supervising the company's business while CH_Founder_2 stayed.

After his time at CH_Company, CH was selected for an entrepreneurship scholarship and founded a startup. With a young team of four people with different backgrounds, he tried to establish a digital education platform. His company created an MVP, secured investments, and hired multiple people for sales, marketing, and software development. After one and a half years, they quit this project due to a lack of perceived market acceptance by potential customers. Afterward, CH remained in the entrepreneurial ecosystem for finance solutions and joined a Southern European startup focusing on a platform for sustainable investing for a half-year project. Later and until today, CH started working for another startup in northern Germany with a product idea similar to the one he tried to create.

Before joining CH_Company, CH did not consider becoming an entrepreneur one day. Even after he started his position there, "it was never within realistic reach to found a venture." Despite some lectures at university, the topic of entrepreneurship never crossed his path. Therefore, CH started to explore the entrepreneurial world with a low initial passion for entrepreneurship.

Considering CH's perceptions of his founders' effort and passion, CH reported different perceptions. On the one hand, he perceived his founders to be low-passionate and to put low effort into their tasks while working for CH_Company. For example, CH described that his founders felt attached to other projects during their time at CH_Company and that he assumes that they are people who always enjoy trying out new projects. Therefore, they did not express or show passion for their current venture but already looked at other projects. Concerning their effort, CH estimated their average working hours for the startup by "far below 40" hours a week because they "did many other things" in parallel. He observed CH_Founder_1 reduce his operative involvement step by step until he only held supervising tasks. He explained his understanding of his supervisors' behavior as they had to split their attention to engage in other projects. On the other hand, he perceived them to be highly passionate about other projects outside CH_Company because they tend to initiate multiple projects in parallel. CH reported that his founders worked on building a platform for entrepreneurs and establishing a venture capitalist firm while they ran CH_Company. CH perceived his founders as "creative people who always want to let off steam in new projects.". He was impressed by the entrepreneur's high passion for founding which impacted him highly.

Perceiving the behavior of his supervisors had a considerable effect on CH as he got inspired by them. CH participated in an entrepreneurship program which CH_Founder_1 did as well in the past, and started his venture afterward, which indicates an increase in entrepreneurial passion. Additionally, he reported that after his time at CH_Company, he developed a high passion for being creative, finding solutions, and creating things himself which are all types of entrepreneurial passion. It is mainly his passion for inventing that increased as he reported that "[...] creating something is something that I really enjoy because it's kind of 'your baby' [laughing]. And be it somehow a product, a website, or anything else." His effort level shows parallels to the behavior of his founders as he put "inconsistent" effort into his venture where "there were probably also weeks where [he] really worked full-time on it, and there were also weeks [...] where [he] worked very little and only a few hours on it.". When I asked what had to be different so he would have spent more effort, he replied, "fewer distractions, fewer side projects. So, it's just that I was doing quite a lot in parallel at the time.". It resembles the perceived behavior of his supervisors. Concerning effort, he also mentioned that perceiving a high level of effort impressed him. However, it also made him feel intimidated, that he needs his work-life balance in working environments and that he is currently not ready to go the effort commitment needed to start another venture again.

Overall, CH started with a low initial passion for entrepreneurship. He perceived low effort and low passion from his founders while working for CH_Company, but simultaneously, he perceived high passion from them while they were working on other projects. Inspired by the founders, his entrepreneurial passion increased while his effort lowered.

4.1.6. The case of JK

JK is a 25-year-old German business computer science graduate who worked for a startup called JK_Company that has been operating in the gastronomy sector for more than four years. He worked in distinct roles as a customer success manager and product manager for the company and was one of the first employees hired. He supported developing an app enabling customers to order food, pay, and collect points. The company had to shut down during the pandemic because most restaurants and cafes were not allowed to operate, resulting in a shortage of money. Consequently, JK left the firm. However, the company was acquired mid-2022 and now runs under new leadership.

Since he was the first employee, he frequently communicated with the founding team. It consisted of three young and inexperienced students who met during an entrepreneurship program at their university. They barely gained working experience outside of internships and working student jobs. During the time of the company's foundation, they had an overall full-time working experience of fewer than three years combined, mainly in consulting. All three came in with different academic backgrounds graduating in business, psychology, and computer science in their master's programs. After his time at JK_Company, JK continued their idea with new developments in his startup. He developed an MVP of an app and hired some employees to grow the venture. Currently, he is looking for investors to expand the firm and gain market share. On social media, he describes himself as an open-minded person willing to go the extra mile.

JK developed a high initial passion for entrepreneurship before joining JK_Company, as he was born and raised in a family of many entrepreneurs. His grandparents and some of his uncles and aunts had owned businesses and showed their passion for entrepreneurship. His family affected him, as he explained: "[...] I think that rubbed off relatively early on because I saw how freely they could move around, how much fun they had [...].". Subsequently, he thought about owning a business himself one day, which he fulfilled later.

While working for JK_Company, JK perceived high entrepreneurial effort and passion from his founders. He reported that they never worked less than 80 hours per week. He described their effort as the "prime example of going the extra mile within those years.". He felt that his founders "only live for this project," which led one to join a retreat because of burnout symptoms. JK highlighted the founders' love for their product, as they might even be the people who used it most. He explained that passion was very influential during his time at JK_Company.

Working for JK_Company increased JK's level of effort. He said that "there were extremely many weeks where I didn't get out of the office with less than 60-70 hours, and I really pushed things through on the weekends because I had the feeling that it was necessary." and that "it has often gone beyond, very often gone beyond" a regular working framework. Additionally, he took on extra tasks and responsibilities to help the firm. For example, he independently invested additional effort in learning in-app design to support the founders in this task, as they could not hire someone for that. Also, his passion for entrepreneurship increased as he reported working with a high passion for JK_Company. He argues that one of the main reasons for his effort was his "passion for this topic." He described that this passion emerged and grew with the increase of his involvement in entrepreneurial tasks. For example, he said, "Personally, I would say that this entrepreneurial passion has definitely increased due to the responsibility I got in the product area.".

While analyzing the interview with JK, it became clear that he was highly affected by his founders as he rooted his decision to start a venture in perceiving his founders. For example, he explained: "That is to say, this bridge, this connection, between what I had with the founders, what was now in retrospect, of course, very unfortunate that it did not work, but was crucial and determining for my decision to say I will now found a company myself.", or "I think maybe the bottom line of why I personally made this decision to say I'm willing to do this myself was actually the - I would call it - the rise and fall of what the founders themselves have gone through.". This effect goes in line with the mentioned contagion mechanisms.

To summarize, JK started to work for JK_Company with high initial entrepreneurial passion and became acquainted with a founding team that he perceived to be highly passionate and high in effort. This perception increased his passion and effort, resulting in him becoming a startup founder.

4.1.7. The case of LL

LL, a holder of a master's degree in finance and management from a German university, is 25 years old and has always wanted to become an entrepreneur. He developed an entrepreneurial mindset early on during school as he started to buy different consumer goods on the internet and sold them to his former classmates, friends, and family. During his bachelor studies, he co-founded a management service for serviced apartment providers in a big German city, which operated for one year but closed due to a lack of demand and the focus on other projects. LL has gained experiences in several European firms, such as banks, consultancies, startups, and venture capitalists. He was also a member of an entrepreneurial network. After finishing the bachelor's program, he got an offer to work for a big consultancy firm. However, he rejected it because it would not be relatable enough to entrepreneurship. Parallel to his master's, he worked for a venture capitalist where he intensively collaborated with early-staged ventures in the entrepreneurial opportunity development processes. Then, after graduating, he tried to start a venture with two colleagues. Later, to establish a business plan and a business case and pitch their ideas to

several investors, they delayed the final commitment to start operating for several reasons. However, the main reason was that they wanted to become more experienced before taking the risk of creating a serious business themselves.

LL finally signed as chief of staff for a construction-tech startup in northern Germany. It is the same company where RB works. RB referred LL for this study. They know each other from entrepreneurial networks. LL started half a year earlier than RB at this company. To stick to the naming convention and for simplicity reasons, I will refer to the firm as LL_Company, although it is the same as the RB_Company. In this role, he acts as the right hand of the founders and collaborates closely with all of them. Additionally, he acted as interim head of people for the company during the time of the interview. LL joined LL_Company with a high initial entrepreneurial passion. As already described, he started to gain interest in entrepreneurship when he was a child, gained several working experiences in the entrepreneurial environment, founded a venture himself, and pursued an entrepreneurial career.

LL perceived low entrepreneurial effort from his founders compared to the other cases. He reported that "they're already stepping on the gas. But I think at a healthy level and could still be a bit more." when I asked about their level of effort. According to him, other entrepreneurs put more effort into their ventures. Additionally, he sometimes feels they want to delay specific tasks or decisions, although, in LL's opinion, they could tackle them the same evening, implicating a lack of perceived effort. Suitably, LL described them as focused on "having a healthy lifestyle" when he talked about the perceived effort and that they suggested he should step down sometimes. Concerning their entrepreneurial passion, he perceived a low level of passion. He described that they do not express their passion "Jordan Belfort-like" because they are somewhat "restrained," and overall, they "don't show their passion that often to the outside." As an example, he described their company meetings where the "speech from them, that doesn't appeal to [him] so much. [They] could make it more emotional.".

While working for LL_Company, LL's level of effort at the starting point was very high, but lately, he questioned his effort level and thought about decreasing hit. Usually, he works "clearly over 40 hours" a week, completes tasks on weekends, and has projects where he worked for multiple weeks "until the middle of the night." He claimed that stepping on the gas even more, would not be possible. In situations where his founders express a low level of effort, he said, "yes, that annoys me extremely because I think I give 110% here, and I want to tear the thing down. [...] And then, of course, you question yourself in one situation or another. Is it worth it? Shouldn't you also slow down a bit?" and "It just gets on your nerves, and you think that they're not there today with the right attitude, and that drags you down. So, I would definitely agree that this influences my attitude.". This led him to question his level of effort or, as he said: "I often ask myself the same question. Whether it would be right to make 'piano,' because they advise me with their experience and say go a little slower.", resulting in actually decrease the effort due to a low perception of effort: "And when you realize, okay, the boss doesn't have the drive or doesn't have the energy, then you think, why should I put in the energy?". When asked about the impact of this perception on his passion for entrepreneurship, his answers diverged. On the one hand, he told that "the fire still burns" and that he still wants to found himself again. On the other hand, he mentioned that "[...] there is the negative impact of 'well, you see how it actually works in reality." and that it is not that easy and "[...] glorified, as it is just always presented from the outside [...]". Subsequently, he concluded about his development of passion: "So I'd have to say it reduced, but on a different basis, because now I have more experience and more knowledge and skills and so on.".

To summarize, LL started his job at LL_Company with a high entrepreneurial passion. He perceived his founders to invest a low level of effort with low passion, which resulted in him questioning his level of effort and reducing his passion for entrepreneurship. In the case of LL, the contagion mechanism became evident in the negative manner in which he described himself by saying: "But of course, if you just think, yes, the founder isn't working too hard, and the company isn't successful either, then that's just frustrating. Then you think, what am I actually doing? If he's not up for it, I'm not up for it either, and then maybe you don't feel like founding a company anymore if you've had such a negative experience.". That statement indicates that his level of effort and passion dropped.

4.1.8. The case of AW

AW is 25 years old and works as a business development representative in the sales area for a startup in south Germany that develops a communication app for other businesses. She holds a degree in fashion management from a business school. AW has several experiences as a working student for industry and fashion firms in HR. However, she had no touching points with entrepreneurship prior to working for AW_Company in her current role. Neither did she ever want to become an entrepreneur herself. As a business development representative, she does not share collaborative tasks with the founding team. However, due to the venture's early stage, she perceives her founders daily at work.

The founding team consists of AW_Founder_1 and AW_Founder_2. AW_Founder_1 is a business management graduate who worked for four years as a project leader for a big German production firm before founding AW_Company. In parallel, he is a member of the advisory board of another venture. On his social media profiles, he claims he is very passionate about the firm and the product he creates. AW_Founder_2 is a serial entrepreneur who graduated in media management and enterprise communications. He worked for several firms for nearly four years before founding his first firm, a web design consultancy. Then, he came together with AW_Founder_1 to create AW_Company.

AW described herself as someone who "honestly didn't have such intensive thoughts" about entrepreneurship before

joining AW_Company, which actually "was a bit random." She mentioned that passion for entrepreneurship was not her "main focus or main reason for taking the job." Therefore, AW joined the startup with a low initial entrepreneurial passion.

AW perceives her supervisors to work with much effort and passion. She described her founders as "crazy working" and that she "would assume that their life is our company. So, there is no separation either spatially or in terms of time. There is no separation between private life and AW_Company, so they live for it and are absorbed in it. It's crazy how much they invest in it.". Additionally, she perceives her founders to be "very passionate," that "they live for it," and that "they really put their heart and soul and their time into it.".

Perceiving her founders working with high effort on their tasks has neither changed AW's effort nor her passion for entrepreneurship. She reported that she works less than other people in her firm, which she is happy about. Besides, she said about the effect of the perceived effort: "It doesn't affect me so much that I now say I have to adapt there; I have to work just as much.". She describes the workload related to founding a firm as a "deterrent" and concludes that a founder would "not have a private life anymore," which would not suit her way of life. Although she was "within a very short time [...] inspired by the construct of a startup" and that she sees "how cool it can be to have your own startup because you can fulfill your dreams," she does not feel the willingness to found a firm one day because of the high amount of effort needed.

All in all, AW started her position at AW_Company with low entrepreneurial passion, which has not changed during her working period until today. AW perceived high effort and passion from her entrepreneurs. According to her, that perception had no noticeable impact on her passion and effort.

Table 3 shows a summary of the findings from the within-case analysis. It captures the initial passion, development of passion and effort, and perceived passion and effort for each case. As GD's perceptions and own developments changed with the dropout of the original founding team, the results altered over time. I marked it in the table with an arrow meaning the first entry represents the state before the dropout, and the second entry represents the state after the dropout. In six out of eight cases, I assigned the same level of perceived effort and perceived passion. It supports the initial assumption of the combined contagion and self-regulation framework that employees concurrently perceive passion while perceiving effort and that these constructs are associated. Therefore, I propose the following:

> *Proposition 1a: High perceived entrepreneurial effort increases the likelihood of perceiving high entrepreneurial passion.*

> *Proposition 1b: Low perceived entrepreneurial effort increases the likelihood of perceiving low entrepreneurial passion.*

4.2. Case patterns and between-case analysis

In the cases of MS, JK, and until the point when the founders left the company in the case of GD, all three cases showed positive synergies concerning entrepreneurial passion while perceiving high effort. The case of LL and the case of GD after the dropout of the founding team showed a negative effect on entrepreneurial passion while perceiving low effort. The mentioned cases align with what I expected as an outcome according to the combined theoretical framework on contagion and self-regulation that explains how perceived effort could affect the employees' passion response. MS, JK, and GD, with her initial founders, reported that they perceived high effort by their supervisors. These perceptions, in turn, triggered an increase in their effort, which the mentioned contagion theory can explain. Then, explained by self-regulation theory, the higher level of effort positively impacted their passion for entrepreneurial tasks. Therefore, working for their corresponding firms made them more passionate about entrepreneurial tasks than they were prior to their employment. The opposite happened in the case of LL, and the case of GD after the new CEOs took over. Both perceived their corresponding supervisors to work on their tasks with little effort. This perception led GD to decrease her effort and LL to develop a tendency to reduce his effort. In turn, their entrepreneurial passion declined as both wanted to become entrepreneurs before their employment, and now they are questioning that plan. In addition, the high-effort cases also showed that the participants perceived high passion from the entrepreneurs, and the low-effort cases showed a low passion perception by the employees. Therefore, perceiving effort might covariate with perceiving passion, with both going in the same direction.

In the cases of RB and AW, the data revealed that the perception of effort did not influence their entrepreneurial passion. The data in the cases of JG and CH showed counterintuitive results according to the temporal theoretical framework. JG perceived high effort, which increased his level of effort but instead of his entrepreneurial passion increasing, it declined. CH perceived a low level of effort by his supervisor that lowered his level of effort, which impacted him even after leaving the firm. Instead of his passion for entrepreneurial tasks declining, his passion increased as he later founded a firm, where, according to him, he failed to succeed due to his low level of effort.

4.2.1. Proximity to the founders, entrepreneurial-relatedness of the tasks, and initial entrepreneurial passion

All cases that went along the expected outcome, speaking of MS, JK, and GD for observing high effort, which increased their entrepreneurial passion, and speaking of LL and GD for observing low effort, which decreased their passion followed a pattern. All these employees have high proximity to their founders, their tasks are highly entrepreneurial-related, and they had high initial entrepreneurial passion before starting the role in their corresponding startup. This finding indicates that a particular involvement of these three factors is a possible requirement for the initiation of the proposed effects of

Table 3: Summary of within-case analysis

Case	Initial Passion	Perceived Effort	Perceived Passion	Effort development	Passion development
JG	Low	High	High	Lower	Lower
GD	High	High → Low	High → Low	Higher → Lower	Higher → Lower
RB	High	Neutral	High	Neutral	Neutral
MS	High	High	High	Higher	Higher
CH	Low	Low	High	Lower	Higher
JK	High	High	High	Higher	Higher
LL	High	Low	Low	Lower	Lower
AW	Low	High	High	Neutral	Neutral

the combined mechanism. In the cases where the mechanism worked as expected, these three factors were above a specific boundary compared to the other cases where the observation did not follow the expected way.

Proximity can be understood broadly in the employee-entrepreneur relationship. It can include the employee's feeling of a close connection to the supervisors, frequent communication, spatial proximity with sitting next to each other in the same office, and a role-based close collaboration as being the head of a particular department, for example. Having high proximity between the entrepreneur and the employee increases the frequency and intensity of the contagion effect as both subjects have more and deeper points of interaction which facilitates the effect's occurrence. It aligns with other scholars' assumptions that employees feel being in the same boat as their founders (Breugst et al., 2012). Higher proximity to the founder at work helps to get richer perceptions of the entrepreneurial effort. It is because the employee perceives the entrepreneur more often, evaluates his level of effort more reliably and draws more accurate inferences about how effort leads to progress for the venture, which is necessary for effort contagion to occur. Table 4 shows examples of statements about proximity to the founders from the interview participants.

Having tasks that are rich in entrepreneurial content is also essential for observing the expected effects. These tasks can typically include acts of inventing or developing the firm. Usually, these tasks have a high degree of decisional freedom, high creativity, deep product involvement, high hierarchy levels, or high responsibility. Working on these types of tasks with the entrepreneur facilitates the contagion effect as these tasks are more similar to the actual tasks of the entrepreneur and, therefore, closer related to entrepreneurial passion. It increases the likelihood for the employee to perceive a goal that is more accessible and eases goal contagion. Putting effort into entrepreneurial-related tasks helps the self-regulation mechanism affect the passion for entrepreneurial tasks. Table 5 displays the comments of the interviewed employees on their tasks. The found patterns in effort perception lead to the development of these propositions:

Proposition 2a: A high proximity to the founders increases the likelihood that perceived effort triggers the combined contagion and self-regulation mechanism, where high perceived effort increases entrepreneurial passion and low perceived effort decreases entrepreneurial passion.

Proposition 2b: A high entrepreneurial-relatedness of the employee's task increases the likelihood that perceived effort triggers the combined contagion and self-regulation mechanism, where high perceived effort increases entrepreneurial passion and low perceived effort decreases entrepreneurial passion.

Proposition 2c: A high initial entrepreneurial passion increases the likelihood that perceived effort triggers the combined contagion and self-regulation mechanism, where high perceived effort increases entrepreneurial passion and low perceived effort decreases entrepreneurial passion.

4.2.2. Effort-Passion discrepancy and the perceived effort-passion antagonism

Suppose proximity to the founders, entrepreneurial-relatedness of the employee's tasks, or initial entrepreneurial passion are low. In that case, this might change the expected effect of perceived effort on the employee's passion response, as in the cases of JG and CH. As shown in the case of JG, he had high proximity to his founders, but his tasks were low in entrepreneurial-relatedness, and he had low initial passion. Concurrently, the temporal model did not deliver the expected outcome. He reported perceiving high entrepreneurial effort by his supervisors, which triggered his effort and increased it. Instead of his entrepreneurial passion increasing like in the other cases, it declined as he did not enjoy the effort, which distanced him from an entrepreneurial career, as he explained. Although the supervisor's effort successfully transferred from JG_Founder_1 to JG, the self-regulation mechanism did not occur. The turning point in his case might be his initial entrepreneurial passion which was relatively low compared to other cases. JG had no touching point with entrepreneurship prior to his role at JG_Company.

Table 4: Interview statements about the degree of proximity

Employee	Subjective degree of proximity	Statement about proximity
JG	High	In fact, there were two of us at one table, which was intended for one person, because the company has grown extremely quickly, and we had to hold the microphones shut. That's how close we were sitting to each other. That means I was able to hear a lot of how she did it, and that's why I can say that she did it very well.
GD	High	Yeah, especially for the technical-based founder. We were working together; we had this normal agile methodology. You have to do daily meetings.
RB	High	Right now, I'm working closely with the founder, or rather I'm in close coordination with him, but it's also because, at the moment, my project is being taken over by the founder.
MS	High	In general, you also have to say MS_Founder_1, the CTO, is super transparent. You can approach him at any time directly via Slack. He will always answer, and he will always take time, no matter what concern you have.
CH	Low	He was always sitting around the corner from me, but somehow, I didn't have that much to do with him. So, we have not chatted so much.
JK	High	So, I was like the fourth or the fifth person in this company.
LL	High	Sure, I also have a super close role with the founders and a great relationship of trust.
AW	Low	Personal contact is rather rare. I don't have any meetings or anything like that. What I did have a month ago, which I thought was really cool, was that our founder, our CEO, set up a meeting for me for a lunch date over two hours, and then we just had a random chat.

The discrepancy between his low initial passion and the high perception and development of effort may prevent the development of positive entrepreneurial passion in a way that this discrepancy has a deterrent effect, lowering the level of entrepreneurial passion. One could say that the demanded effort was too big compared to his passion for entrepreneurship. Metaphorically spoken, too much firewood is bad for a fire, so too much high perceived effort may be bad for passion, "the fire of desire" (Cardon et al., 2009, p. 515).

CH had low proximity, low entrepreneurial-relatedness in his tasks, and low initial passion. He perceived low effort by his supervisors, which decreased his effort. Instead of his entrepreneurial passion declining, it increased to a level where he started a venture and became an entrepreneur. However, his founders were highly involved in other venture creation processes, making CH very passionate about founding. Consequently, a high discrepancy between CH's effort and the perceived entrepreneurial passion emerged. Therefore, perceiving high passion might overcome the negative effect of perceiving low effort if the discrepancy between low effort and high perceived passion is big enough. It indicates that perceiving high passion might impact the employees' passion response more than perceiving low effort. Spoken in metaphors again, when trying to make a fire, the low perceived effort might resemble wet firewood, but perceived passion might resemble the gasoline. The low perceived supervisor effort infected him through effort contagion, and he reduced his effort. However, the positive passion transfer through emotional contagion had a more significant effect on his entrepreneurial passion. Overall, using the combined mechanism of contagion and self-regulation, I did not expect the outcome of the two mentioned cases. These findings lead to the development of the following propositions:

> *Proposition 3: A high discrepancy between initial entrepreneurial passion and perceived effort triggers a decrease in entrepreneurial passion.*
>
> *Proposition 4: A high discrepancy between high perceived passion and decreased effort triggers an increase in entrepreneurial passion.*
>
> *Proposition 5: Perceived passion has a stronger effect on entrepreneurial passion than perceived effort.*

4.2.3. No influence under certain conditions

The cases of RB and AW showed that under certain conditions, it might be that perceiving effort does not influence the employee's behavior and emotion at all. In the case of RB, the employee had high proximity to the entrepreneurs, his tasks were rich in entrepreneurial content, and he had a high initial passion for entrepreneurship. However, he reported perceiving a neutral level of entrepreneurial effort. Consequently, there is no observable change in the data concerning his effort or passion for entrepreneurial tasks. It induces that neither a contagion nor a self-regulation effect occurred when RB perceived neither a low nor a high level of effort from his supervisors. Therefore, there must be a boundary in the perceived effort level in both directions, low and high, to be crossed so that effort contagion occurs.

Table 5: Interview statements about the degree of entrepreneurial-relatedness of the tasks

Employee	Subjective degree of entrepreneurial-relatedness of tasks	Statement about entrepreneurial-relatedness of tasks
JG	Low	I worked as a bit of a jack-of-all-trades because we were deep in customer acquisition, and I was partly in sales.
GD	High	I mean, I always wanted to see the whole startup cycle. From product creation to programming to bringing to customers and startup companies are mainly giving you that option. So it was cool for me to have conversations with the customers and sometimes go to onsite visits and help them instead of just programming in a closed room.
RB	High	There are founder topics that we are working on, strategic topics.
MS	High	All of this actually gives me a lot of entrepreneurial freedom, but I also have to say, I think it's also a special environment because it's just super encouraged, and if you take responsibility on your own, that's more than welcome.
CH	Low	I think I would rather look at experiences outside of CH_Company because at CH_Company; I was rather less entrepreneurially active myself.
JK	High	I played a major role in determining where the product goes.
LL	High	I've been there since the beginning of the year as Chief of Staff. Which is ultimately the role in which you are kind of like the right hand of the founders. You are also the sparring partner or strategic resource and can then always act like a kind of firefighter then in the individual teams and departments to help there and to work with them then also with the management there.
AW	Low	And my job is to maintain the initial contact. Bringing all the leads into the sales process. As I said, all of my tasks involve acquiring the first customer and then passing them on to our account executives, who then follow the rest of the sales process.

In the case of AW, the employee perceived high entrepreneurial effort and passion. However, the data did not show that this impacted her effort or her passion for entrepreneurial tasks. AW has low proximity to her founders and low entrepreneurial-relatedness in her tasks. Compared to all other cases, she had the lowest initial entrepreneurial passion as she never had a touching point with entrepreneurship besides one course at university. She never thought or showed interest in someday becoming an entrepreneur or owning a business. It induces a low initial passion for entrepreneurship, besides having low proximity to the founders and low entrepreneurial-relatedness of the tasks, which did not initiate the combined contagion and self-regulation mechanism as the contagion effect could not be triggered. One might say it is not about "be there or be square"; instead, an employee must have a certain degree of entrepreneurial passion, proximity to the founders, and entrepreneurial-related tasks so that the perception of effort can affect the employee's entrepreneurial passion.

As the cases of RB and AW showed no influence on perceiving entrepreneurial effort under certain conditions, I make the following propositions:

Proposition 6: Neutral perceived entrepreneurial effort does not affect entrepreneurial passion.

Proposition 7: A combination of low proximity to the founders, low entrepreneurial-task relatedness, and low initial entrepreneurial passion does not affect entrepreneurial passion.

The findings and the derived propositions are summarized and included in the model represented in Figure 3. I based this model on combined theory from literature and enriched it with the results of this study. Therefore, I further developed the theory and set new theoretical boundaries.

5. Discussion

I motivated this research by the limited available knowledge in the field of entrepreneurship on the impact of perceived behavior on experienced emotions. To extend the current empirical works, I researched the question: "How does the employees' perception of their supervisor's effort influence the employees' entrepreneurial passion?". The data revealed that proximity to the founders, entrepreneurial-relatedness of the employee's tasks, and initial entrepreneurial passion work as antecedents of the combined mechanism of contagion and self-regulation. No effect will likely occur if they are not present to a certain degree. Furthermore, perceived neutral effort showed no effect on entrepreneurial passion. Additionally, the data showed that counterintuitive results might occur while perceiving effort. Perceiving high effort under the condition of having a low initial passion can lead to a decrease in

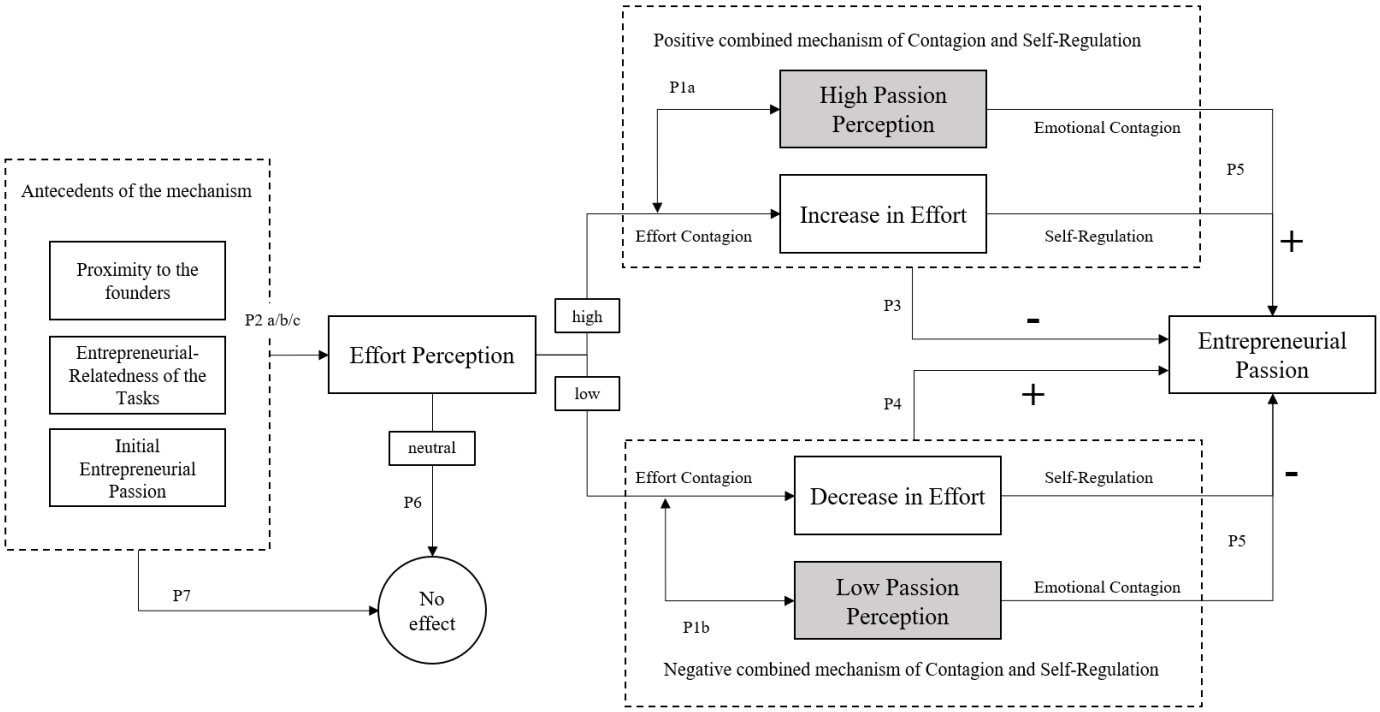


Figure 3: A model of perceived supervisor effort and its impact on the employee's entrepreneurial passion

entrepreneurial passion. In contrast, despite perceiving low effort, entrepreneurial passion can increase under perceiving high passion and having low initial entrepreneurial passion.

5.1. Theoretical and practical implications

My finding on perceived effort as an indirect antecedent of entrepreneurial passion joins the literature stream that deviates from the dominant theories in entrepreneurship (e.g., Gielnik et al. 2015) that view passion as a catalyst for effort (e.g., Baum and Locke 2004; Cardon et al. 2009). The result follows the call from Gielnik et al. (2015, p. 1025) "[...] that theoretical frameworks on entrepreneurial passion need to consider entrepreneurial passion to be an outcome of entrepreneurial effort." and extend it by including perceived effort and the entrepreneurial employee into this framework.

Additionally, this paper responds to calls in the entrepreneurship literature to look into what fuels the passion for entrepreneurship (Cardon et al., 2012) by accounting for perceived entrepreneurial effort as an essential factor. Research in this field is relevant since academics suggested that entrepreneurial passion is positively related to crucial characteristics in entrepreneurship, for example, creativity (Cardon, 2008), success and commitment (Breugst et al., 2012), motivation (Cardon et al., 2005), evaluations (Mitteness et al., 2012) and numerous more.

Besides, these findings answer the calls from academics to investigate more on the impact of entrepreneurs on new venture employees (Breugst et al., 2012; Hubner et al., 2020) and therefore contribute to leadership research which has focused dominantly on leadership styles and organizational performance (Ensley et al., 2006). Scholars argued for the importance of passionate employees in new ventures to be beneficial (Cardon, 2008) and positively impacts their motivation, creativity, and success at work (Ho & Pollack, 2014). Cardon (2008) asked how founders may shift their passion to their employees while assuming the positive benefits of having passionate employees. The results give another valid answer to this question by implying that perceiving effort can transfer passion from entrepreneur to employee. As Hubner et al. (2020) argued on the importance of stimulating the employees' entrepreneurial passion as this can be a relevant strategy for maximizing their contributions to the entrepreneurial goal, this paper shows another pathway by examining how perceived effort triggers an employee's passion response. Additionally, this paper supports the claim of Cardon (2008, p. 83): "[...] if entrepreneurs want their employees to experience passion they must work harder in order to make their passion contagious to their employees." in explaining how and why working harder makes entrepreneurial passion contagious.

By examining the effect of perceived effort on the employees' entrepreneurial passion, this work is first in combining theory on passion contagion, goal contagion, and self-regulation and therefore contributes to their streams of literature. The result suggests that it is not either the perception of behavior or the perception of emotion to stimulate the emotion of entrepreneurial passion exclusively but that both mechanisms run hand in hand. Employees cannot per-

ceive emotion without perceiving behavior, nor can they perceive behavior while not perceiving emotion. This paper indicated that both streams work collaboratively and should be observed as a whole system concurrently.

Finally, the results show that the positive and negative effects of perceiving effort have certain boundaries. For example, on the one hand, employees can become less passionate about entrepreneurship when they experience too intense effort contagion in relation to their initial entrepreneurial passion, which can make them feel deterrent about entrepreneurship. On the other hand, perceiving high entrepreneurial passion can outshine the effects of perceiving low entrepreneurial effort, especially when the employee's initial passion is low.

The results hold several implications for practitioners. First, entrepreneurs can stimulate their employees' emotions by showing specific behavior. Entrepreneurs who work with high effort on their entrepreneurial tasks are likely to observe an increase in their employees' passion for entrepreneurship when they maintain high proximity to them and provide them with entrepreneurial tasks. Showing high effort can benefit entrepreneurs as their employees' efforts will likely increase and deliver better performance. Employees who experienced the positive effects of the combined contagion and self-regulation mechanism can improve venture performance by contributing with higher creativity, higher task performance, more innovation, and even intrapreneurship. However, entrepreneurs must be careful when employees perceive low entrepreneurial effort, as this could negatively impact the employees' level of effort and their entrepreneurial passion. Entrepreneurs should aim to counter this mechanism by expressing high entrepreneurial passion or, as suggested by Lex et al. (2019), encouraging low-passionate employees to increase their efforts to make them more passionate. Besides, they should keep an eye on not overwhelming low-passionate employees with entrepreneurial effort and passion. This perception may decrease the employees' passion response because it can make them feel a deterrent and more distanced from entrepreneurial tasks.

Employees can include these findings when choosing to work for startups. Suppose they want to increase their entrepreneurial passion. In that case, they could emphasize their potential supervisors' level of entrepreneurial effort and aim for roles with high founder proximity and high entrepreneurial-relatedness of the tasks. Employees should know they can counter a decrease in their effort and passion when perceiving low effort by raising their effort or looking out for highly passionate supervisors.

5.2. Limitations, future research, and conclusion

A potential limitation of this study reveals due to its case-based nature. The case-based methodology of this study restricts the model's generalizability and, therefore, its external validity because of the limited sample size. This limitation is usual for case-based approaches and qualitative studies in general (Lee et al., 1999). However, it provides an intriguing opportunity for further and future research. To evaluate key linkages in the presented model, scholars could use broader methodologies to generalize the findings beyond the case-based approach. For instance, they could take already established measures for constructs like effort (Foo et al., 2009), passion (Cardon et al., 2013), established adaptions to scales to account for the employees' perceptions of these constructs like perceived passion (Breugst et al., 2012) or formulate own adaptions for perceived effort. Furthermore, they should build measures for proximity to the founder and entrepreneurial-relatedness of the employees' tasks. With these measures, future scientists can test the presented propositions quantitatively.

Another area for improvement in this study is that it could be more dynamic. Although I tried to account for the development in passion and effort by posing questions to the interview participants and selected participants who can describe current processes and past results, this study does not investigate a phenomenon over time. As Gielnik et al. (2015) observed, current literature trends consider that people's motivation and self-regulation are not static and change over time. Future contributions should account for that and establish longitudinal studies. Time is essential, especially in process work, as it is the only omnipresent factor (Gehman et al., 2018). A suggested benchmark would be to measure the proposed key linkages every week, as other scholars demonstrated before (e.g., Gielnik et al. 2015).

It is possible to argue that this study suffers from obtrusiveness that impacts the participants' self-reported measures. Obtrusiveness is a known issue in qualitative studies (Lee et al., 1999). I countered obtrusiveness by establishing a high level of researcher-subject trust through theoretical sampling, as I knew most of the participants before or got at least a warm introduction. Future research could improve by relying on more objective measurements for constructs like passion and effort. In this study, I evaluated key linkages through the self-reported measures of the interview participants. I tried to validate the employees' statements by triangulating them with secondary data to enrich the cases, but this does not substitute more objective data like video recordings at work or captured timetables.

This research combined the literature streams investigating perceived emotion and perceived behavior. This work shows that both mechanism, effort contagion, and passion contagion can occur, and the data suggests that perceiving passion has a stronger effect on entrepreneurial passion than perceiving effort. Future research should answer why one can be stronger than the other and what factors contribute to this.

Finally, an exciting path for further investigation could be on the upper boundary for perceiving effort on this model. The findings imply a certain maximum of perceived effort compared to the initial entrepreneurial passion the employee brings into the job. The perceived effort might negatively affect the employee's entrepreneurial passion when this boundary is exceeded. It is relevant and interesting for scholars and practitioners to discover further insights into this relationship.

To conclude, the perception of effort is a relevant topic in entrepreneurship as every employee in new ventures is exposed to perceive entrepreneurial effort. This study shows that employees are affected by the mere perception of effort in their passion for entrepreneurship. When initial passion, proximity to the founders, and entrepreneurial-relatedness of the employees' tasks are high, the perception of supervisor effort can have a noticeable impact on the employee. Then, employees who perceive high effort are likely to increase their effort, which in turn triggers an increase in their entrepreneurial passion. On the other hand, employees who perceive low effort might reduce their effort, which triggers a decrease in entrepreneurial passion. Furthermore, perceiving too much effort could lead to a decline in entrepreneurial passion if the employee initially has a low passion for entrepreneurship. Entrepreneurs and employees should account for this. Besides, the perception of low effort can get outshined by perceiving high passion. It is another major result of this study as it implies that perceived emotion has a bigger impact on the recipient's emotion than perceived behavior.

References

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology*, *87*(1), 23–37. https://doi.org/10.1037/0022-3514.87.1.23

Aguinis, H., & Solarino, A. M. (2019). Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal*, *40*(8), 1291–1315. https://doi.org/10.1002/smj.3015

Bandura, A. (1988). Organisational Applications of Social Cognitive Theory. *Australian Journal of Management*, *13*(2), 275–302. https://doi.org/10.1177/031289628801300210

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*, 1–26. https://doi.org/10.1146/annurev.psych.52.1.1

Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-Efficacy: The Exercise of Control. *Journal of Cognitive Psychotherapy*, *13*(2), 158–166. https://doi.org/10.1891/0889-8391.13.2.158

Baron, R. A. (2008). The Role of Affect in the Entrepreneurial Process. *Academy of Management Review*, *33*(2), 328–340. https://doi.org/10.5465/amr.2008.31193166

Barsade, S. G. (2002). The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. *Administrative Science Quarterly*, *47*(4), 644–675. https://doi.org/10.2307/3094912

Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, *89*(4), 587–598. https://doi.org/10.1037/0021-9010.89.4.587

Baum, J. R., Locke, E. A., & Smith, K. G. (2001). A Multidimensional Model of Venture Growth. *Academy of Management Journal*, *44*(2), 292–303. https://doi.org/10.5465/3069456

Baumeister, R. F., Vohs, K. D., DeWall, C. N., & Zhang, L. (2007). How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc*, *11*(2), 167–203. https://doi.org/10.1177/1088868307301033

Bingham, C. B., & Eisenhardt, K. M. (2011). Rational heuristics: the 'simple rules' that strategists learn from process experience. *Strategic Management Journal*, *32*(13), 1437–1464. https://doi.org/10.1002/smj.965

Bitektine, A. (2008). Prospective Case Study Design. *Organizational Research Methods*, *11*(1), 160–180. https://doi.org/10.1177/1094428106292900

Breugst, N., Domurath, A., Patzelt, H., & Klaukien, A. (2012). Perceptions of Entrepreneurial Passion and Employees' Commitment to Entrepreneurial Ventures. *Entrepreneurship Theory and Practice*, *36*(1), 171–192. https://doi.org/10.1111/j.1540-6520.2011.00491.x

Breugst, N., Patzelt, H., & Rathgeber, P. (2015). How should we divide the pie? Equity distribution and its impact on entrepreneurial teams. *Journal of Business Venturing*, *30*(1), 66–94. https://doi.org/10.1016/j.jbusvent.2014.07.006

Breugst, N., Patzelt, H., & Shepherd, D. A. (2020). When is Effort Contagious in New Venture Management Teams? Understanding the Contingencies of Social Motivation Theory. *Journal of Management Studies*, *57*(8), 1556–1588. https://doi.org/10.1111/joms.12546

Brundin, E., Patzelt, H., & Shepherd, D. A. (2008). Managers' emotional displays and employees' willingness to act entrepreneurially. *Journal of Business Venturing*, *23*(2), 221–243. https://doi.org/10.1016/j.jbusvent.2006.10.009

Cardon, M. S. (2008). Is passion contagious? The transference of entrepreneurial passion to employees. *Human Resource Management Review*, *18*(2), 77–86. https://doi.org/10.1016/j.hrmr.2008.04.001

Cardon, M. S., Foo, M.-D., Shepherd, D., & Wiklund, J. (2012). Exploring the Heart: Entrepreneurial Emotion is a Hot Topic. *Entrepreneurship Theory and Practice*, *36*(1), 1–10. https://doi.org/10.1111/j.1540-6520.2011.00501.x

Cardon, M. S., Glauser, M., & Murnieks, C. Y. (2017). Passion for what? Expanding the domains of entrepreneurial passion. *Journal of Business Venturing Insights*, *8*, 24–32. https://doi.org/10.1016/j.jbvi.2017.05.004

Cardon, M. S., Gregoire, D. A., Stevens, C. E., & Patel, P. C. (2013). Measuring entrepreneurial passion: Conceptual foundations and scale validation. *Journal of Business Venturing*, *28*(3), 373–396. https://doi.org/10.1016/j.jbusvent.2012.03.003

Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2005). Entrepreneurial Passion: The nature of emotions in entrepreneurship. *Academy of Management Proceedings*, *2005*(1), G1–G6. https://doi.org/10.5465/ambpp.2005.18778641

Cardon, M. S., Wincent, J., Singh, J., & Drnovsek, M. (2009). The nature and experience of entrepreneurial passion. *Academy of Management Review*, *34*(3), 511–532. https://doi.org/10.5465/amr.2009.40633190

Carver, C. S. (2006). Approach, Avoidance, and the Self-Regulation of Affect and Action. *Motivation and Emotion*, *30*(2), 105–110. https://doi.org/10.1007/s11031-006-9044-7

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Chen, X., Yao, X., & Kotha, S. (2009). Entrepreneur Passion And Preparedness In Business Plan Presentations: A Persuasion Analysis Of Venture Capitalists' Funding Decisions. *Academy of Management Journal*, *52*(1), 199–214. https://doi.org/10.5465/amj.2009.36462018

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, *13*(1), 3–21. https://doi.org/10.1007/BF00988593

Corcoran, K., Brohmer, H., Eckerstorfer, L. V., & Macher, S. (2020). When your goals inspire my goals: The role of effort, personal value, and inference in goal contagion. *Comprehensive Results in Social Psychology*, *4*(1), 78–108. https://doi.org/10.1080/23743603.2020.1767502

Dik, G., & Aarts, H. (2007). Behavioral cues to others' motivation and goal pursuits: The perception of effort facilitates goal inference and contagion. *Journal of Experimental Social Psychology*, *43*(5), 727–737. https://doi.org/10.1016/j.jesp.2006.09.002

Dik, G., & Aarts, H. (2008). I Want to Know What You Want: How Effort Perception Facilitates the Motivation to Infer Another's Goal. *Social Cognition*, *26*(6), 737–754. https://doi.org/10.1521/soco.2008.26.6.737

Drnovsek, M., Cardon, M. S., & Patel, P. C. (2016). Direct and Indirect Effects of Passion on Growing Technology Ventures. *Strategic En-*

trepreneurship Journal, *10*(2), 194–213. https://doi.org/10.1002/sej.1213

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of Management Review*, *32*(4), 1155–1179. https://doi.org/10.5465/amr.2007.26586086

Edwards, R., & Holland, J. (2013). *What is Qualitative Interviewing? 'What is?' Research Methods Series*. Bloomsbury Academic.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, *14*(4), 532–550. https://doi.org/10.5465/amr.1989.4308385

Eisenhardt, K. M. (2021). What is the Eisenhardt Method, really? *Strategic Organization*, *19*(1), 147–160. https://doi.org/10.1177/1476127020982866

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32. https://doi.org/10.5465/amj.2007.24160888

Eitam, B., & Higgins, E. T. (2010). Motivation in mental accessibility: Relevance Of A Representation (ROAR) as a new framework. *Social and Personality Psychology Compass*, *4*(10), 951–967. https://doi.org/10.1111/j.1751-9004.2010.00309.x

Elfenbein, H. A. (2014). The many faces of emotional contagion. *Organizational Psychology Review*, *4*(4), 326–362. https://doi.org/10.1177/2041386614542889

Ensley, M. D., Hmieleski, K. M., & Pearce, C. L. (2006). The importance of vertical and shared leadership within new venture top management teams: Implications for the performance of startups. *The Leadership Quarterly*, *17*(3), 217–231. https://doi.org/10.1016/j.leaqua.2006.02.002

Epstude, K., & Mussweiler, T. (2009). What you feel is how you compare: How comparisons influence the social induction of affect. *Emotion*, *9*(1), 1–14. https://doi.org/10.1037/a0014148

Foo, M.-D., Uy, M. A., & Baron, R. A. (2009). How do feelings influence effort? An empirical study of entrepreneurs' affect and venture effort. *Journal of Applied Psychology*, *94*(4), 1086–1094. https://doi.org/10.1037/a0015599

Gatewood, E. J., Shaver, K. G., Powers, J. B., & Gartner, W. B. (2002). Entrepreneurial Expectancy, Task Effort, and Performance. *Entrepreneurship Theory and Practice*, *27*(2), 187–206. https://doi.org/10.1111/1540-8520.00006

Geen, R. G. (1991). Social motivation. *Annual Review of Psychology*, *42*, 377–399. https://doi.org/10.1146/annurev.ps.42.020191.002113

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2018). Finding Theory–Method Fit: A Comparison of Three Qualitative Approaches to Theory Building. *Journal of Management Inquiry*, *27*(3), 284–300. https://doi.org/10.1177/1056492617706029

Gielnik, M. M., Spitzmuller, M., Schmitt, A., Klemann, D. K., & Frese, M. (2015). "I Put in Effort, Therefore I Am Passionate": Investigating the Path from Effort to Passion in Entrepreneurship. *Academy of Management Journal*, *58*(4), 1012–1031. https://doi.org/10.5465/amj.2011.0727

Hatfield, E., Cacioppo, J. T., & Rapson, R. L. (1993). Emotional Contagion. *Current Directions in Psychological Science*, *2*(3), 96–100. https://doi.org/10.1111/1467-8721.ep10770953

Hayton, J. C. (2003). Strategic human capital management in SMEs: An empirical study of entrepreneurial performance. *Human Resource Management*, *42*(4), 375–391. https://doi.org/10.1002/hrm.10096

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley & Sons Inc. https://doi.org/10.1037/10628-000

Helt, M. S., Fein, D. A., & Vargas, J. E. (2020). Emotional contagion in children with autism spectrum disorder varies with stimulus familiarity and task instructions. *Development and Psychopathology*, *32*(1), 383–393. https://doi.org/10.1017/S0954579419000154

Hitt, M. A. (2000). The new frontier: Transformation of management for the new millennium. *Organizational Dynamics*, *28*(3), 7–17. https://doi.org/10.1016/S0090-2616(00)88446-6

Ho, V. T., & Pollack, J. M. (2014). Passion Isn't Always a Good Thing: Examining Entrepreneurs' Network Centrality and Financial Performance with a Dualistic Model of Passion. *Journal of Management Studies*, *51*(3), 433–459. https://doi.org/10.1111/joms.12062

Huber, G. P., & Power, D. J. (1985). Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. *Strategic Management Journal*, *6*(2), 171–180. https://doi.org/10.1002/smj.4250060206

Hubner, S., Baum, M., & Frese, M. (2020). Contagion of Entrepreneurial Passion: Effects on Employee Outcomes. *Entrepreneurship Theory and Practice*, *44*(6), 1112–1140. https://doi.org/10.1177/1042258719883995

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly*, *24*(4), 602–611.

Kruger, J., Wirtz, D., van Boven, L., & Altermatt, T. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, *40*(1), 91–98. https://doi.org/10.1016/S0022-1031(03)00065-9

Laurin, K. (2016). Interpersonal influences on goals: Current and future directions for goal contagion research. *Social and Personality Psychology Compass*, *10*(11), 668–678. https://doi.org/10.1111/spc3.12289

Lee, T. W., Mitchell, T. R., & Sablynski, C. J. (1999). Qualitative Research in Organizational and Vocational Psychology, 1979–1999. *Journal of Vocational Behavior*, *55*(2), 161–187. https://doi.org/10.1006/jvbe.1999.1707

Lex, M., Gielnik, M. M., & Frese, M. (2019). Effort and success as predictors of passion. In *Passion for Work: Theory, Research, and Applications* (pp. 227–259). Oxford University Press.

Lex, M., Gielnik, M. M., Spitzmuller, M., Jacob, G. H., & Frese, M. (2020). How Passion in Entrepreneurship Develops Over Time: A Self-Regulation Perspective. *Entrepreneurship Theory and Practice*, *46*(4), 985–1018. https://doi.org/10.1177/1042258720929894

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705–717. https://doi.org/10.1037/0003-066x.57.9.705

Locke, E. A., & Latham, G. P. (2006). New Directions in Goal-Setting Theory. *Current Directions in Psychological Science*, *15*(5), 265–268. https://doi.org/10.1111/j.1467-8721.2006.00449.x

Loersch, C., Aarts, H., Keith Payne, B., & Jefferis, V. E. (2008). The influence of social groups on goal contagion. *Journal of Experimental Social Psychology*, *44*(6), 1555–1558. https://doi.org/10.1016/j.jesp.2008.07.009

Loersch, C., & Payne, B. K. (2014). Situated Inferences and the What, Who, and Where of Priming. *Social Cognition*, *32*(Supplement), 137–151. https://doi.org/10.1521/soco.2014.32.supp.137

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

Mitteness, C., Sudek, R., & Cardon, M. S. (2012). Angel investor characteristics that determine whether perceived passion leads to higher evaluations of funding potential. *Journal of Business Venturing*, *27*(5), 592–606. https://doi.org/10.1016/j.jbusvent.2011.11.003

Moriano, J. A., Molero, F., Topa, G., & Lévy Mangin, J.-P. (2014). The influence of transformational leadership and organizational identification on intrapreneurship. *International Entrepreneurship and Management Journal*, *10*(1), 103–119. https://doi.org/10.1007/s11365-011-0196-x

Murnieks, C. Y., Mosakowski, E., & Cardon, M. S. (2014). Pathways of Passion. *Journal of Management*, *40*(6), 1583–1606. https://doi.org/10.1177/0149206311433855

Palomares, N. A. (2013). When and How Goals Are Contagious in Social Interaction. *Human Communication Research*, *39*(1), 74–100. https://doi.org/10.1111/j.1468-2958.2012.01439.x

Philippe, F. L., Vallerand, R. J., Houlfort, N., Lavigne, G. L., & Donahue, E. G. (2010). Passion for an activity and quality of interpersonal relationships: The mediating role of emotions. *Journal of Personality and Social Psychology*, *98*(6), 917–932. https://doi.org/10.1037/a0018017

Platow, M. J., Haslam, S. A., Both, A., Chew, I., Cuddon, M., Goharpey, N., et al. (2005). "It's not funny if they're laughing": Self-categorization, social influence, and responses to canned laughter. *Journal of Experimental Social Psychology*, *41*(5), 542–550. https://doi.org/10.1016/j.jesp.2004.09.005

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of Management*, *12*(4), 531–544.

Pratt, M. G. (2009). From the Editors: For the Lack of a Boilerplate: Tips on Writing Up (and Reviewing) Qualitative Research. *Academy of Management Journal*, *52*(5), 856–862. https://doi.org/10.5465/amj.2009.44632557

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*(1), 145–172. https://doi.org/10.1037/0033-295x.110.1.145

Schoenewolf, G. (1990). Emotional contagion: Behavioral induction in individuals and groups. *Modern Psychoanalysis*, *15*(1), 49–61.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques*. Sage Publications Thousand Oaks, CA.

Suddaby, R. (2006). From the Editors: What Grounded Theory is Not. *Academy of Management Journal*, *49*(4), 633–642. https://doi.org/10.5465/amj.2006.22083020

Sullins, E. S. (1991). Emotional Contagion Revisited: Effects of Social Comparison and Expressive Style on Mood Convergence. *Personality and Social Psychology Bulletin*, *17*(2), 166–174. https://doi.org/10.1177/014616729101700208

Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Leonard, M., & Marsolais, J. (2003). Les passions de l'ame: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, *85*(4), 756–767. https://doi.org/10.1037/0022-3514.85.4.756

Visser, V. A., van Knippenberg, D., van Kleef, G. A., & Wisse, B. (2013). How leader displays of happiness and sadness influence follower performance: Emotional contagion and creative versus analytical performance. *The Leadership Quarterly*, *24*(1), 172–188. https://doi.org/10.1016/j.leaqua.2012.09.003

Weick, K. E. (1984). Small wins: Redefining the scale of social problems. *American Psychologist*, *39*(1), 40–49. https://doi.org/10.1037/0003-066x.39.1.40

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Motivationen und Resultate der Anti-Konsum Praxis ‚Stooping'

Motivations and Outcomes of the Anti-Consumption Practice 'Stooping'

Leonie Böhm

Freie Universität Berlin

Abstract

The climate crisis of the 21st century demands new and sustainable forms of alternative consumption. Within the scope of this bachelor thesis, the emerging anti-consumption practice of "Stooping", a form of second-hand consumption where items are placed on the street by consumers and can be freely taken and used by others, is examined for the first time. The aim of this paper is to explore possible motivations and outcomes of Stooping. These aspects are extracted from the existing literature on second-hand and anti-consumption and subsequently investigated through qualitative interviews. The results indicate that economic motives, hedonistic values, the simplicity of consumption decisions, sustainability considerations, the upcycling of found products, and the good quality of used products can serve as motivations for Stooping. Outcomes include an increased connection between the seller and the buyer, an increased connection between the buyer and the product, as well as a positive social perception of the consumers. Investigating new consumption trends like Stooping contributes to a better understanding of sustainable consumption practices and can help position them as conscious and deliberate alternatives to traditional consumption.

Zusammenfassung

Die Klimakrise des 21. Jahrhunderts fordert neue und nachhaltige Formen des alternativen Konsums. Im Rahmen dieser Bachelorarbeit wird die aufstrebende Anti-Konsum Praxis Stooping, eine Art des Second-Hand Konsums, bei dem Gegenstände von Konsumierenden an die Straße gestellt und von anderen kostenfrei konsumiert werden können, erstmals untersucht. Das Ziel dieser Arbeit ist, mögliche Motivationen und Resultate des Stooping zu erforschen. Diese werden aus der existierenden Literatur zu Second-Hand- und Anti-Konsum herausgearbeitet und anschließend mithilfe qualitativer Interviews überprüft. Die Ergebnisse zeigen, dass ökonomische Motive, hedonistische Werte, die Einfachheit der Konsumentscheidungen, Nachhaltigkeitsüberlegungen, das Aufwerten von gefundenen Produkten und eine gute Qualität gebrauchter Produkte Motivationen für Stooping darstellen können. Resultate des Stooping können eine verstärkte Verbundenheit zwischen Verkäufer*in und Käufer*in, zum Objekt sowie eine positive soziale Wahrnehmung der Konsumierenden sein. Die Untersuchung neuer Konsumtrends wie Stooping führt zu einem besseren Verständnis nachhaltiger Konsumpraktiken und kann dazu beitragen, diese als bewusste und geplante Konsum-Alternative zu positionieren.

Keywords: alternative consumption; anti-consumption; second-hand consumption; stooping

Gerne möchte ich mich bei Catalina Wache und Prof. Dr. Jana Möller-Herm für die Ermöglichung dieses interessanten Bachelorarbeits-Themas bedanken. Über ein Thema zu schreiben, dass mich auch privat bewegt, hat mir sehr viel bedeutet. Vielen Dank auch an Catalina für die Unterstützung während des Schreibprozesses und die Möglichkeit, Probleme jederzeit anbringen zu können. Zuletzt möchte ich mich bei allen Freunden und Verwandten bedanken, die mir viel Inspiration und immer ehrliches Feedback für die Arbeit gegeben haben.

DOI: https://doi.org/10.5282/jums/v9i3pp1665-1680

1. Einleitung

In unserer konsumorientierten Gesellschaft hat sich in den letzten Jahren der Fokus immer mehr auf das Thema Nachhaltigkeit gerichtet. Als Reaktion darauf entwickelten sich einige alternative Formen des Konsums, wie der Anti-Konsum oder der Konsum gebrauchter Produkte. Vor allem der Erwerb und die Nutzung von gebrauchten Produkten stellen einen wichtigen Konsum-Trend dar, der sich in letzter Zeit immer weiter verbreitet hat (z. B. momox AG, 2022). Immer mehr Menschen werden sich den Auswirkungen ihres Konsums bewusst und versuchen, auf eine nachhaltige und verantwortungsvolle Weise zu konsumieren. Dabei kaufen nicht nur Personen mit einem niedrigen Einkommen Second-Hand Produkte wie Kleidung, auch Personen, die mehr Geld zur Verfügung haben, konsumieren gebraucht (momox AG, 2022). Neben ökonomischen und nachhaltigen Motiven gibt es noch viele weitere Gründe, Second-Hand Produkte neuen Produkten vorzuziehen (Pavlakoudis, 2022). Doch nicht nur Second-Hand Shopping in klassischen Kanälen, wie auf dem Flohmarkt oder im Second-Hand Geschäft wird immer beliebter. Auch andere Formen des Konsums gebrauchter Produkte, wie das Tauschen von Produkten oder das Verkaufen und Verschenken aussortierter Produkte in der Nachbarschaft werden nach und nach von der breiten Bevölkerung angenommen. Eine dieser neuen Formen des Konsums gebrauchter Produkte ist Stooping. Hierbei werden nicht mehr benötigte Produkte an die Straße gestellt, so dass andere Personen diese ohne monetäre Gegenleistung mitnehmen können. Stooping lässt sich als Form des Anti-Konsums und des Second-Hand Konsums einordnen, hat aber die Besonderheit, dass der Konsum gratis und häufig innerhalb der Nachbarschaft der Konsumierenden stattfindet. Der Begriff Stooping kommt nicht aus der Wissenschaft, sondern wurde durch diverse Nutzer*innen auf Plattformen in den sozialen Medien geprägt (z. B. TheSorryGirls, 2022). Zu den verbreiteteren Formen des alternativen Konsums Second-Hand Konsum und Anti-Konsum existiert bereits viel Literatur. Zu der neueren Form Stooping, die sich aktuell in vielen dicht besiedelten, städtischen Regionen verbreitet, gibt es jedoch wenig Forschung. Aus diesem Grund beschäftigt sich diese Arbeit näher mit Stooping unter Zuhilfenahme der bereits bestehenden Literatur zu Second-Hand Konsum und Anti-Konsum. Genauer sollen die Motivationen der Personen, die gebrauchte Produkte von der Straße konsumieren, beleuchtet werden. Außerdem soll herausgefunden werden, was die Resultate eines solchen Konsums sein können. Nach den Erläuterungen verschiedener Formen des alternativen Konsums und der Einordnung von Stooping als Anti-Konsum Praxis und als Second-Hand Konsum, werden mögliche Motivationen und Resultate von Stooping mithilfe der Literatur herausgearbeitet. Einige Motivationen und Resultate lassen sich aus bereits vorhandenen Erkenntnissen zu Second-Hand Konsum und Anti-Konsum ableiten, andere werden aus weiteren Fachgebieten hergeleitet. Im Anschluss wird das methodische Vorgehen bei der qualitativen Untersuchung in Form von Interviews erklärt und die Untersuchungsergebnisse dargestellt. Zuletzt folgt eine Diskussion der Ergebnisse, inklusive der Limitationen und des weiteren Forschungsbedarfs.

2. Theoretische Grundlagen

Um ein grundlegendes Verständnis für die theoretischen Grundlagen zu erhalten, werden in diesem Kapitel drei Konzepte des alternativen Konsums aufgezeigt und definiert. Außerdem werden mithilfe der Literatur mögliche Motivationen und Resultate von Stooping erarbeitet.

2.1. Konzepte des alternativen Konsums

2.1.1. Anti-Konsum

Der Begriff Konsum bezeichnet alle Prozesse des Erwerbs, der Nutzung und Entsorgung von Gütern durch Individuen (Lee et al., 2011; McCracken, 1990). Güter können nicht nur Produkte sein, sondern auch Ideen, Dienstleistungen, Marken oder Erfahrungen (Lee et al., 2011).

Alternativer Konsum untersucht die Beweggründe einer Person, keine traditionellen Güter über traditionelle Kanäle zu konsumieren (Close & Zinkhan, 2009; Gould et al., 1997). Eine Form dieses alternativen Konsums ist Anti-Konsum. Anti-Konsum beschäftigt sich mit den Gründen, die gegen Konsum beziehungsweise gegen den Erwerb, die Nutzung und Entsorgung von Gütern, sprechen (Chatzidakis & Lee, 2013; Lee et al., 2011). Es geht also um das Gegenteil von Konsum. Dabei muss jedoch gesagt werden, dass die Gründe *gegen* Konsum nicht immer automatisch gleichzusetzen sind mit dem Gegenteil der Gründe *für* Konsum. So sind z. B. nicht alle Konsument*innen, die keine nachhaltigen Produkte kaufen, automatisch gegen eine nachhaltige Lebensweise (Chatzidakis & Lee, 2013). Anti-Konsum kann sich als Widerstand oder Abneigung gegenüber Konsum widerspiegeln (Zavestoski, 2002). Es können bestimmte Marken, Produkte, Dienstleistungen oder die Konsum-Kultur im Allgemeinen abgelehnt werden (Chatzidakis & Lee, 2013). Damit ein Prozess als Anti-Konsum Praxis eingeordnet werden kann, ist es wichtig, dass dieser Prozess absichtlich und bewusst ausgeführt wird (Chatzidakis & Lee, 2013; Cherrier et al., 2011; García-de-Frutos et al., 2018). So ist z. B. das Nicht-Kaufen eines Produkts aus dem Grund, dass ein Produkt einer anderen Marke vorgezogen wird, nicht als Anti-Konsum zu definieren (Cherrier et al., 2011). Auch das Nicht-Kaufen eines Produktes, dessen Konsum vorher gar nicht erst in Betracht gezogen wurde, gilt nicht als Anti-Konsum (García-de-Frutos et al., 2018). Zwar richtet sich Anti-Konsum per Definition gegen Konsum, aber ein konsequenter Anti-Konsum ist schwer zu erreichen und auch nicht erstrebenswert (Luchs et al., 2011). Es bedarf eines Minimums an Konsum, um in der modernen Gesellschaft zu bestehen (Albinsson & Yasanthi Perera, 2012) und eine soziale Akzeptanz zu erreichen (den sogenannten „glass floor“) (Cherrier et al., 2012; García-de-Frutos et al., 2018). Albinsson und Yasanthi Perera (2012) stellen zudem fest, dass Konsumierende, die sich als Anti-Konsument*innen identifizieren, eher für Anti-Überkonsum

oder Anti-Konsum von Produkten einstehen, die schädlich für das persönliche und soziale Wohlergehen sind.

Lee et al. (2011) teilen Anti-Konsum außerdem in drei Bereiche ein: Das *Ablehnen* von gewissen Gütern (reject), das *Einschränken* des Konsums bestimmter Güter (restrict) und das *ideologische Umdenken* bezüglich des Prozesses des Erwerbs, der Benutzung und der Entsorgung von Objekten (reclaim). Ein Beispiel für das Umdenken des klassischen Konsums ist die Praxis des ‚Mülltauchens'. Hierbei durchforsten Personen Müllcontainer auf der Suche nach brauchbaren Dingen. So geben sie den Dingen, die für andere Leute Müll sind, für sich eine neue Bedeutung (Lee et al., 2011).

2.1.2. Second-Hand Konsum

Second-Hand Shopping bezeichnet den Erwerb von gebrauchten Produkten durch Shopping-Kanäle, die sich von solchen für neue Produkte unterscheiden (Guiot & Roux, 2010; Luchs et al., 2011; Steffen, 2017). Second-Hand Produkte können Kleidungsstücke (Xu et al., 2014; Zahid et al., 2022), aber auch Möbelstücke oder Bücher sein. Analog zur Definition von Konsum, stellt Second-Hand Konsum also die Nutzung von gebrauchten Produkten dar, und den Erwerb und die Entsorgung dieser über Second-Hand Kanäle. Solche Kanäle können Wohltätigkeitsorganisationen, Auktionen, Online-Auktionen (z. B. eBay), Verkäuferseiten, Vintage- oder Second-Hand Geschäfte sowie Flohmärkte sein (Charbonneau, 2008; Waight, 2013). Durch eine spezielle Art der Entsorgung von Produkten, kann Second-Hand Konsum, im Gegensatz zum herkömmlichen linearen Konsum neuer Produkte, als Kreislauf stattfinden (Luchs et al., 2011). So ist der Konsum eines Produktes nicht bereits mit der Entsorgung des Produktes beendet. Vielmehr wird durch eine spezielle Art der Entsorgung der erneute Erwerb und Nutzung des Objektes möglich (Luchs et al., 2011; Xu et al., 2014). Diese Arten der Entsorgung können das Spenden, das Schenken und das Tauschen von Produkten sein (Luchs et al., 2011), auch z. B. innerhalb der Familie oder unter Freunden (Charbonneau, 2008). Da beim Konsum von Second-Hand Produkten also keine traditionellen Güter über traditionelle Kanäle konsumiert werden, kann Second-Hand Konsum als Konzept des alternativen Konsums angesehen werden.

Second-Hand Konsum stellt ein ideologisches Umdenken bezüglich der Prozesse des Erwerbs, der Nutzung und der Entsorgung von Produkten (reclaim nach Lee et al., 2011) dar, ist also auch eine Art von Anti-Konsum. Allerdings nur unter der Bedingung, dass Second-Hand Konsum bewusst und absichtlich dem Konsum neuer Produkte vorgezogen wird (Chatzidakis & Lee, 2013; Cherrier et al., 2011; García-de-Frutos et al., 2018). Personen, die Second-Hand Konsum betreiben, besitzen Gründe, die gegen herkömmlichen Konsum sprechen und weichen aus diesem Grund auf Second-Hand Kanäle aus. Black und Cherrier (2010) berichten außerdem von Handlungen der Ablehnung, der Reduzierung und des Wiederverwendens als Praktiken des Anti-Konsums in Bezug auf Nachhaltigkeit. Auch Second-Hand Konsum ist eine Art des Wiederverwendens von Produkten. Second-Hand Konsum ist vielseitig, so können nicht nur neuwertige gebrauchte Produkte über Second-Hand Kanäle erworben werden, sondern auch Produkte, die wieder ihren Zweck erfüllen, wenn sie repariert oder aufgewertet werden (Guiot & Roux, 2010; Luchs et al., 2011). Diese Reparaturen können ein Gefühl von Befriedigung bei den reparierenden Personen auslösen (Guiot & Roux, 2010). Und auch das Nutzen eines Produktes für einen neuen Zweck stellt Second-Hand Konsum dar (Charbonneau, 2008).

Second-Hand Konsum hat eine lange Tradition in Europa (Guiot & Roux, 2010). In Nord-England wurde z. B. schon während des 18. Jahrhunderts gebrauchte Kleidung vertrieben und konsumiert (Lambert, 2004). Und auch in den letzten Jahren gab es einen klaren Anstieg von Second-Hand Konsum. Das Marktvolumen von Second-Hand Kleidung in den USA lag 2017 noch bei 20 Milliarden US-Dollar, 2021 waren es schon 35 Milliarden US-Dollar. Prognostiziert werden für 2026 sogar 82 Milliarden US-Dollar (thredUP, 2022). Und während bei einer Umfrage in Deutschland 2020 noch 56% der Befragten angaben, schon einmal Second-Hand Kleidung gekauft zu haben, waren es 2021 bereits 67% der Befragten (momox AG, 2022). Der Konsum von Second-Hand Produkten verlängert das Leben bereits existierender Produkte und spart somit wertvolle Ressourcen ein (Luchs et al., 2011). Dieser Aspekt ist vor allem bezüglich der aktuellen klimatischen Entwicklungen von Vorteil. So können z. B. durch den Kauf von Second-Hand Kleidung im Vergleich zu herkömmlicher neuer Mode mehr als 239 kg CO_2 pro Kleidungsstück eingespart werden (zum Vergleich: Beim Kauf von nachhaltiger Kleidung sind es 105kg CO_2 pro Kleidungsstück) (thredUP, 2020).

2.1.3. Stooping

Stooping bedeutet wortwörtlich übersetzt sich zu beugen oder zu bücken (DeepL SE, n. d.). Doch aktuell wird dieser Begriff außerdem verwendet (überwiegend in den USA und Kanada), um einen Konsum-Trend zu beschreiben. Bei diesem Konsum-Trend geht es darum, Produkte, wie Kleidungsstücke, Möbel oder Bücher, die andere Personen an die Straße gestellt haben, zu entdecken und sie mit nach Hause zu nehmen. Bei Stooping handelt es sich um einen Begriff, der nicht aus der Wissenschaft kommt und der somit wissenschaftlich nicht exakt definiert ist. Vielmehr findet sich der Begriff auf diversen Plattformen in den sozialen Medien wieder. Zum Beispiel beschreiben „TheSorryGirls" in einem ihrer Videos auf YouTube Stooping als das Beschaffen von Objekten, die an die Straße gestellt wurden, um als Sperrmüll abgeholt zu werden oder in der Hoffnung, dass andere Personen diese mitnehmen (TheSorryGirls, 2022). Werden Produkte mit der Absicht an die Straße gestellt, dass sie von anderen Personen konsumiert werden können, befinden sich meistens ‚zu verschenken'- oder ‚zum Mitnehmen'-Kennzeichnungen an ihnen. Außerdem gibt es einige Plattformen in den sozialen Medien wie Instagram oder Telegram, die Stooping für Städte organisieren. So kann z. B. ein Produkt auf der Straße entdeckt, ein Bild an die Plattform geschickt und das Produkt dadurch auch von anderen Personen entdeckt werden (stoopingberlin, n. d.; StoopingNYC, n. d.). Da das Konzept

am besten funktioniert, wenn viele Leute beieinander wohnen, die häufig ihre Sachen aussortieren und an die Straße stellen, ist es vor allem ein städtisches Phänomen. Zum aktuellen Zeitpunkt gibt es wenig bis keine Forschung zu diesem Themengebiet.

Da Stooping eine Art des Konsums ist, gibt es auch hier drei Prozesse: Den Erwerb, die Nutzung und die Entsorgung. Bezüglich des *Erwerbs* wird bei Stooping nicht neu konsumiert, sondern gebraucht. Produkte können ohne Gegenleistung von der Straße mitgenommen werden. Bei der *Nutzung* werden Produkte aufgrund existierender Funktion oder potenzieller Reparatur nicht weggeworfen, sondern können erneut in den Konsum-Kreislauf eingeführt werden. So können Produkte noch effizienter genutzt werden (Luchs et al., 2011). Die *Entsorgung* findet beim Stooping so statt, dass nicht mehr gebrauchte Produkte an die Straße gestellt werden. Genau wie beim Second-Hand Konsum kann auch Stooping nur stattfinden, wenn der Konsum als Kreislauf gesehen wird. Gegenstände können nur durch Stooping erworben werden, wenn die vorherigen Besitzer*innen sie angemessen auf der Straße bereitstellen. Diese Art der Entsorgung stellt sich aber als kompliziert dar. Es ist nicht immer klar, ob Produkte mit der Intention an die Straße gestellt wurden, weggeschmissen zu werden oder von anderen Personen mitgenommen zu werden. Aus diesem Grund beschäftigen sich die nachfolgenden Kapitel nur mit dem Erwerb und der Nutzung durch Stooping, um den Rahmen dieser Arbeit nicht zu sprengen. Da sich der Erwerb, die Nutzung und die Entsorgung durch Stooping vom Konsum herkömmlicher Produkte unterscheiden, kann Stooping als Konzept des alternativen Konsums angesehen werden. Bei den drei Prozessen des Konsums findet sich bei Stooping auch das ideologische Umdenken dieser wieder (reclaim nach Lee et al. (2011)). Wie oben als Beispiel erwähnt, ist das Durchsuchen des ‚Mülls' anderer Personen und dem Geben einer neuen Bedeutung der Gegenstände genau das, was Personen, die Stooping betreiben, machen. Zusammen mit dem Fakt, dass Stooping außerdem das Wiederverwenden von Produkten darstellt (Black & Cherrier, 2010), sprechen diese zwei Argumente für Stooping als Anti-Konsum Praxis. Des Weiteren werden beim Stooping gebrauchte Produkte über Kanäle konsumiert, die verschieden sind von solchen für neue Produkte. Aufgrund dessen ist Stooping auch als eine Art von Second-Hand Konsum einzuordnen. Die Besonderheit im Vergleich zu klassischem Second-Hand Konsum ist jedoch, dass der Konsum gratis stattfindet. Der Gratiserwerb hat Ähnlichkeiten zum traditionellen Konzept des Schenkens (Belk, 2010; Luchs et al., 2011), mit dem Unterschied, dass sich schenkende und beschenkte Person nicht persönlich kennen und es sich um kein für die Person ausgesuchtes ‚Geschenk' handelt. Außerdem hat Stooping eine Ähnlichkeit zur Form des Leihens beziehungsweise Teilens, die von Luchs et al. (2011) oder Belk (2010) beschrieben worden ist. Hierbei werden Produkte, die sich im eigenen Besitz befinden, von anderen Personen kurzfristig (mit)genutzt und gehen danach wieder in den Besitz zurück. Dieser Konsum findet ebenfalls ohne den Austausch von Geld statt und hat einen starken Gemeinschafts-Gedanken (Luchs et al., 2011). Allerdings werden Produkte bei Stooping nicht nur kurzfristig konsumiert, sondern wechseln dauerhaft den*die Besitzer*in.

Stooping kann am besten als eine soziale Innovation beschrieben werden, also eine Initiative, die von der Gesellschaft ausgeht. Soziale Innovationen sind Phänomene, die sich auf die Entwicklung und den Ausbau alternativer sozialer Praktiken konzentrieren. Diese alternativen Praktiken heben sich maßgeblich von den traditionellen Routinen der Gesellschaft ab. Beispiele sind Plattformen zum Teilen von Produkten oder das eigenständige Herstellen und Aufwerten von Produkten, auch ‚Do-it-Yourself' (DIY) genannt, um dem Massenkonsum entgegenzuwirken (Jaeger-Erben et al., 2015). Bei Stooping werden Produkte ohne Erwartung auf eine Gegenleistung an die Straße gestellt und gratis erworben. Durch Initiativen, wie die oben genannten Instagram-Seiten, kann außerdem eine Plattform entstehen, mit deren Hilfe Produkte gezielt konsumiert werden können. Konsumierte Produkte können außerdem nach dem Erwerb aufgewertet oder repariert werden. Zusammenfassend stellt Stooping also eine neuartige soziale Innovation dar, die als eine Art von Second-Hand Konsum und damit auch als Anti-Konsum einzuordnen ist.

2.2. Mögliche Motivationen und Resultate von Stooping

Laut Guiot und Roux (2010) sind Motivationen für Second-Hand Shopping die psychologischen und materiellen Motive, die Personen dazu bringen, gebrauchte Produkte über Second-Hand Kanäle zu konsumieren. Analog dazu, können die Motivationen für Stooping als psychologische und materielle Motive beschrieben werden, gebrauchte Produkte durch Stooping zu konsumieren. Wolf und McQuitty (2013) arbeiten außerdem neben den Motivationen für ihr beschriebenes Phänomen (DIY-Verhalten) auch verschiedene Resultate des Phänomens aus. Diese Vorgehensweise wird in dieser Arbeit genutzt, um die Motivationen von Stooping in Kapitel 2.2.1 und die Resultate von Stooping in Kapitel 2.2.2 herauszuarbeiten. Hierbei muss jedoch erwähnt werden, dass die Resultate von Stooping auch Motivationen für einen erneuten Konsum-Prozess über Stooping sein können. Wie oben erwähnt beschränkt sich diese Arbeit aus Kapazitätsgründen im Folgenden nur auf die Motivationen, Produkte durch Stooping zu erwerben und die Resultate der Nutzung der erworbenen Produkte. Eine Übersicht der Motivationen und Resultate inklusive der Darstellung, dass sich die Resultate auf die Motivationen auswirken können, befindet sich am Ende des Kapitels (Abbildung 1).

2.2.1. Mögliche Motivationen von Stooping

Beim Erwerb von gebrauchter Kleidung ist es so, dass der geringe Preis ein wesentlicher Grund für Konsumierende ist, diese den herkömmlichen neuen Produkten vorzuziehen (Xu et al., 2014). Über Second-Hand Kanäle ist es außerdem möglich, außergewöhnliche Kleidungsstücke günstiger zu erwerben als in speziellen Boutiquen (Charbonneau, 2008; Reiley & DeLong, 2011) und ‚Schnäppchen zu ergattern' (Xu

et al., 2014). Charbonneau (2008) kam zu der Erkenntnis, dass insbesondere Frauen durch die Möglichkeit qualitative und außergewöhnliche Kleidung zu Schnäppchenpreisen zu bekommen motiviert sind, Second-Hand Kleidung zu kaufen und zu nutzen. Und auch bei anderen Produktkategorien sind gebrauchte Objekte typischerweise günstiger als neue. Die Möglichkeit, Produkte zu einem fairen Preis zu bekommen und weniger Kompromisse in Bezug auf den Erwerb eingehen zu müssen, können Motivationen für Second-Hand Shopping sein (Guiot & Roux, 2010; Williams & Windebank, 2002). Da Stooping eine Art von Second-Hand Konsum ist, könnte auch hier der ökonomische Wert eine Motivation sein, Produkte zu erwerben und zu nutzen. Bei Stooping besteht sogar die Möglichkeit, Produkte gratis zu erwerben und nicht nur zu einem günstigeren Preis, wie es bei traditionellen Second-Hand Kanälen der Fall ist.

> **Proposition 1:** Personen, die Stooping betreiben, werden durch ökonomische Motive angetrieben, also dadurch Geld zu sparen oder Schnäppchen zu machen.

Nicht nur der geringe Preis motiviert Personen Second-Hand zu shoppen, auch das Finden von etwas sehr Wertvollem zu einem verhältnismäßig niedrigen Preis kann beim Erwerb von Second-Hand Kleidung Freude bereiten (Charbonneau, 2008; Weil, 1999; Xu et al., 2014). Auch beim Second-Hand Shopping anderer Produktkategorien kann eine hedonistische Motivation die Schatzsuche nach bestimmten wertvollen und originellen Produkten sein (Guiot & Roux, 2010). Hedonistische Werte sind z. B. Spaß oder Freude (Wolf & McQuitty, 2013). Der hedonistische Wert der Schatzsuche kann also besser als der Spaß beim Streben nach etwas Unerwartetem beschrieben werden, bei dem Glück und Überraschung eine große Rolle spielen (Bardhi & Arnould, 2005). Charbonneau (2008) schreibt von einem Nervenkitzel, den Konsumierende bei der ‚Jagd' nach Produkten empfinden. Die Überraschung und Ungewissheit beim Erwerb von gebrauchten Produkten ergeben sich dadurch, dass man vor dem Erwerb nicht wissen kann, ob ein bestimmtes Produkt gerade im Geschäft verfügbar ist, wie es beim Konsum von neuen Produkten der Fall ist. Abgesehen von Second-Hand Konsum gab es dieses Phänomen auch in der ehemaligen DDR (Deutsche Demokratische Republik), wo ebenso eine Produkt-Knappheit herrschte, was für Spaß der Konsumierenden bei der ‚Schatzsuche' nach gewissen Produkten führte (Albinsson et al., 2010). Bei Stooping herrscht genauso eine Produkt-Knappheit und Ungewissheit. Produkte stehen in einmaliger Ausführung an der Straße und Konsumierende können nicht überprüfen, wann und wo es die Produkte gibt (eine Ausnahme sind die in Kapitel 2.1.3 erwähnten Plattformen in den sozialen Medien, doch auch auf diesen haben Konsumierende keine Garantie, dass das Produkt noch an der Straße steht). Aus diesen Gründen kann die Vermutung aufgestellt werden, dass beim Erwerb durch Stooping hedonistische Werte eine Motivation darstellen.

> **Proposition 2:** Personen, die Stooping betreiben, werden durch hedonistische Werte wie Spaß oder Freude motiviert.

Neben den ökonomischen Motiven und den hedonistischen Werten können auch ökologische Bedenken eine Motivation sein, gebrauchte Produkte (Kleidung) über Second-Hand Kanäle zu erwerben (Charbonneau, 2008; Guiot & Roux, 2010; Reiley & DeLong, 2011). Das Wiederverwenden von noch funktionalen Produkten, die Verminderung der Erschöpfung natürlicher Ressourcen und die Vermeidung der unnötigen Verbreitung weiterer Produkte, können Konsumierende dazu bringen, Second-Hand zu kaufen (Guiot & Roux, 2010). Waight (2013) argumentiert, dass die direkte Wiederverwendung als nachhaltigste Form des Konsums gilt, noch vor Recycling oder der Verwendung von grüner Technologie. Stooping kann bezogen auf Nachhaltigkeit als umweltorientierte Anti-Konsum Praxis angesehen werden. Der umweltorientierte Anti-Konsum besteht aus Aktivitäten, die sich gegen Konsum richten (Reduzieren, Vermeiden, Ablehnen (García-de-Frutos et al., 2018) oder Wiederverwenden (Black & Cherrier, 2010)), um die Umwelt zu schützen (García-de-Frutos et al., 2018). Bei dieser Art von prosozialem Verhalten sind Konsumierende davon überzeugt, dass ein bedeutungsvolles Leben sowohl für einen selbst als auch für die Umwelt und die Gesellschaft gut sein sollte (Balderjahn et al., 2021; De Young, 1996). Stooping ist eine umweltorientierte Anti-Konsum Praxis, wenn es als eine Alternative zum Konsum von neuen Produkten über herkömmliche Kanäle ausgeführt wird (García-de-Frutos et al., 2018). Weil Personen, die umweltorientierte Anti-Konsum Praktiken durchführen, eher im Einklang mit ihren Prinzipien handeln, können diese Praktiken das Wohlbefinden der Personen verbessern (García-de-Frutos et al., 2018; Szmigin et al., 2009). Auch bei Stooping können sich Personen durch den Einklang mit ihren nachhaltig orientierten Prinzipien wohler fühlen als beim herkömmlichen Konsum. Produkte werden gebraucht erworben und wiederverwendet, so dass natürliche Ressourcen geschont werden können.

> **Proposition 3:** Personen, die Stooping betreiben, werden durch Nachhaltigkeitsaspekte motiviert.

2.2.2. Mögliche Resultate von Stooping

Im letzten Kapitel wurden mögliche Motivationen beschrieben, Produkte durch Stooping zu erwerben. In diesem Kapitel sollen die Resultate herausgearbeitet werden, die mit dem Konsum gebrauchter Produkte über Stooping einhergehen könnten. Die Resultate und ein damit einhergehender Wert treten ein, wenn Personen ein Produkt durch Stooping erworben haben und beginnen, dieses zu nutzen.

Beim Stooping interagieren Personen mehr mit ihrem sozialen Umfeld als beim klassischen Konsum neuer Produkte. Der Erwerb der Produkte findet nämlich nicht in einer speziell eingerichteten Institution (Geschäft) statt, sondern überwiegend in der Nachbarschaft der Konsumierenden. Stoo-

ping kann als eine Art von kollaborativem Konsum angesehen werden, bei dem eine oder mehrere Personen Produkte oder Dienstleistungen konsumieren, indem sie mit einer oder mehreren anderen Personen an gemeinsamen Aktivitäten teilnehmen (Felson & Spaeth, 1978), nämlich dem An-die-Straße-Stellen und Sammeln von Objekten. Albinsson und Yasanthi Perera (2012) nennen Umverteilungsmärkte als ein Beispiel für kollaborativen Konsum. Bei diesen können Produkte über Internetplattformen gratis weitergegeben werden. Bei Stooping werden auf eine ähnliche Art und Weise Produkte gratis weitergegeben. Beim unentgeltlichen kollaborativen Konsum kann ein Gefühl der Gemeinschaft eine Motivation sein, an den gemeinsamen Aktivitäten teilzunehmen, aber auch ein Resultat dieser Aktivitäten sein (Albinsson & Yasanthi Perera, 2012). Gemeinschaften haben viele Vorteile, wie Gesundheit, Freude oder verminderte Fälle von sozialen Bedrohungen wie Verbrechen oder Armut (Albinsson & Yasanthi Perera, 2012; Putnam, 2000). Aus diesen Gründen lässt sich also vermuten, dass Stooping als eine Form des kollaborativen Konsums das Gefühl der Gemeinschaft der Personen verstärkt, die an dieser Aktivität teilnehmen.

> **Proposition 4:** Personen, die Stooping betreiben, empfinden ein verstärktes Gefühl der Gemeinschaft.

Neben dem verstärkten Gefühl der Gemeinschaft könnte es bei Personen, die Stooping betreiben auch zu einem verstärkten Gefühl der Verbundenheit zum erworbenen Produkt kommen. Konsumierende besitzen und nutzen Objekte nämlich aufgrund eines Wertes, den sie ihnen geben. Dieser Wert basiert auf der Bedeutung, die den Objekten zugeschrieben wird (Richins, 1994). Bedeutungen können unter anderem die Repräsentation der Beziehungen der Besitzenden zu anderen Personen sein (Kleine et al., 1995; Richins, 1994). So könnte ein Produkt, dass durch Stooping erworben wurde, z. B. mit der Person verbunden werden, mit der man das Produkt entdeckt hat. Eine weitere Bedeutung kann einem Objekt durch die Darstellung der eigenen Identität und dem Ausdruck des Selbst zugeschrieben werden (Belk, 1988; Black & Cherrier, 2010; Kleine et al., 1995; Richins, 1994). Objekte werden noch stärker als Erweiterung des Selbst angesehen, wenn sie bearbeitet werden, da Energie in sie investiert wird (Belk, 1988). Das könnte vor allem bei Stooping der Fall sein, da hier häufig Produkte erworben werden, die noch aufgewertet werden müssen. Neben der Erweiterung des Selbst und der Repräsentation von Beziehungen kann ein Objekt außerdem mit der eigenen persönlichen Geschichte (Richins, 1994) und Erfahrungen (Belk, 1988; Richins, 1994), Leistungen (Belk, 1988), Erinnerungen (Belk, 1988; Kleine et al., 1995; Richins, 1994) oder Gefühlen verknüpft werden (Belk, 1988). Nach Kleine et al. (1995) können vor allem die Darstellung der eigenen Identität und der Beziehungen eines Objekts zu einer Verbundenheit zum Objekt führen. Eine Person kann sich also besonders mit einem durch Stooping erworbenem Produkt verbunden fühlen, da viele und außergewöhnliche Verknüpfungen (Gefühle, Erinnerungen etc.) mit dem Objekt in Verbindung gebracht werden.

> **Proposition 5:** Personen fühlen sich mit Produkten mehr verbunden, wenn sie durch Stooping erworben wurden.

Während es sich bei der Verbundenheit zum Objekt um die Beziehung zwischen dem*der Besitzer*in und dem Objekt handelt, geht es im folgenden Abschnitt um die Beziehung zwischen dem Objekt, dem*der Besitzer*in und dem sozialen Umfeld der Person. Beziehungen mit Objekten sind nicht nur zweiseitig, sondern dreiseitig (Person A – Objekt - Person B) (Belk, 1988). Da Objekte dazu dienen können, die eigene Identität zum Ausdruck zu bringen (Belk, 1988; Black & Cherrier, 2010; Kleine et al., 1995; Richins, 1994), werden diese Darstellungen auch an das soziale Umfeld der Person vermittelt, die das Objekt besitzt. Besitztümer können von der Gesellschaft wahrgenommen und in einem gewissen Kontext interpretiert werden (Richins, 1994). Da Objekte also immer von anderen Personen wahrgenommen werden, können die Konsumentscheidungen einer Person durch die Gesellschaft und das unmittelbare soziale Umfeld beeinflusst werden (Richins, 1994; Waight, 2013). Durch die Vermittlung von Informationen können Objekte außerdem dazu dienen, dass sich Personen besonders und einzigartig fühlen. Zum Beispiel bewirken Vintage-Produkte oder eine geringe Anzahl an nachhaltigen Produkten ein ‚Sich-Abheben' von der breiten Masse (Bellezza, 2022). Ein spezielles Aussehen eines Objektes (z. B. Originalität) gibt dem Objekt also einen höheren Wert (Richins, 1994). Ein Resultat von Stooping könnte dementsprechend sein, dass Personen durch den Erwerb von einzigartigen, originellen Produkten positiv von ihrem sozialen Umfeld wahrgenommen werden.

> **Proposition 6:** Personen, die Stooping betreiben, werden positiv von ihrem sozialen Umfeld wahrgenommen.

3. Methodisches Vorgehen

Um die Motivationen und Resultate von Stooping als Anti-Konsum Praxis besser zu verstehen, wurde eine qualitative empirische Untersuchung durchgeführt. Die Daten wurden gesammelt, um die oben gestellten Propositionen und das aufgestellte Modell zu überprüfen. Außerdem soll herausgefunden werden, ob Konsumierende von weiteren Aspekten motiviert werden, Stooping zu betreiben oder andere Resultate mit der Praxis einhergehen. In den folgenden Kapiteln wird das Forschungsdesign beschrieben, der Aufbau und die Durchführung der Studie genauer erläutert und die erhobenen Daten ausgewertet.

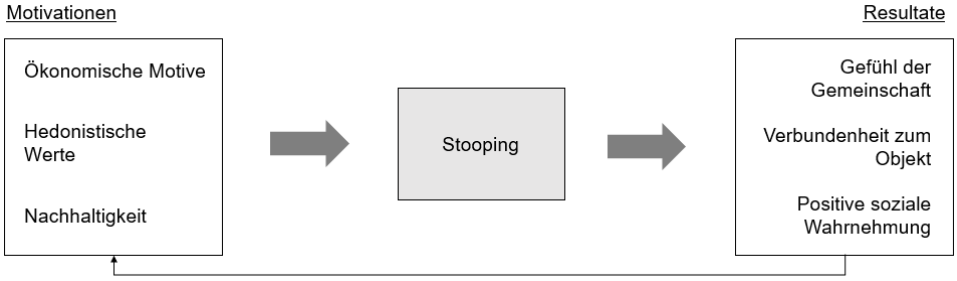


Abbildung 1: Übersicht der möglichen Motivationen und Resultate von Stooping (eigene Darstellung)

3.1. Datenerhebung

3.1.1. Leitfadengestütztes Interview

Da Stooping ein bisher wenig erforschtes Themengebiet darstellt, wurde ein leitfadengestütztes Interview für eine qualitative Untersuchung ausgewählt. Das Ziel einer qualitativen Studie besteht darin, das subjektive Verständnis, die Bedeutung oder den Sinnbildungsprozess von Personen oder Gruppen zu untersuchen (Cassell, 2015). Es wurde ein semi-strukturiertes Interview gewählt, bei dem ein Leitfaden mit Fragen vorliegt. Von diesem Leitfaden kann jedoch abgewichen werden, um interessante Antworten der Interviewpartner*innen bei Bedarf weiter vertiefen zu können (Cassell, 2015). Semi-strukturierte Interviews eignen sich gut, um spezielle Sachverhalte tiefer zu ergründen oder das Verhalten der Interviewpartner*innen in speziellen Situationen besser nachvollziehen zu können (Cassell, 2015). Mit semi-strukturierten Interviews können also sowohl das Konsumverhalten von Personen, die Produkte durch Stooping erwerben, als auch die Motivationen und Resultate dieses Konsums gut untersucht werden. Die insgesamt acht Interviewpartner*innen wurden einzeln interviewt. Einen Überblick über die interviewten Personen gibt Anhang A. Die Namen wurden geändert, um die Anonymität zu wahren.

3.1.2. Aufbau und Durchführung des Interviews

Für die qualitative Studie wurden acht Personen befragt, um eine hinreichend große Menge an Informationen zu erlangen, die Auswertung der Daten dennoch handhabbar zu halten. Nach acht Interviews häuften sich die Antworten zudem, was darauf hindeutete, dass keine neuen Erkenntnisse mehr zu gewinnen waren. Der Interview-Leitfaden wurde mithilfe der erarbeiteten theoretischen Grundlagen und Propositionen erstellt und anhand eines Probeinterviews getestet (Roulston, 2014). Er bestand aus mehreren Blöcken. Zuerst wurden die Interviewpartner*innen nach ihren demografischen Daten befragt. Dann folgten offene Fragen zum allgemeinen Konsumverhalten mit dem Ziel, das Konsumverhalten der Befragten genauer beschreiben zu können und relevante Produkteigenschaften zu ergründen. Anschließend folgte ein Fragen-Block zum Konsumverhalten von gebrauchten Produkten. Dieser Block inkludierte Fragen zum Konsumverhalten von gebrauchten Produkten in der Kindheit der Befragten, dem aktuellen Konsumverhalten gebrauchter Produkte (Zeitpunkt des Konsums, Anlass des Konsums, Kanäle für den Erwerb und die Entsorgung gebrauchter Produkte) und dem Konsumverhalten der Interviewten bei einem dringenden Bedarf. Danach wurden die Interviewpartner*innen zum Thema Stooping befragt. Hierzu wurden die Personen in eine Situation hineinversetzt, in der sie ein Produkt auf der Straße entdecken und es durch Stooping erwerben könnten. Sie wurden gefragt, wie sie auf das Produkt hätten aufmerksam werden können, was ihre Gedanken und Gefühle in dieser Situation wären und wie sich ihr potenzielles Verhalten und weiteres Vorgehen gestalten würden. Außerdem sollten sie erklären, welche Produkte sie am Ehesten über Stooping konsumieren würden. Anschließend wurde der Hauptteil des Interviews geführt. Dieser legte den Schwerpunkt auf die Motivationen und Resultate, Stooping zu betreiben. Hierbei wurden die Personen zuerst direkt nach ihren Motivationen, Stooping zu betreiben, und nach den Resultaten dieses Konsums befragt. Bei Bedarf wurden genannte Motivationen oder Resultate mit Anschlussfragen vertieft. Der Interview-Leitfaden schloss mit Fragen zur Kenntnis des Begriffs ‚Stooping' und zur potenziellen Änderung der Einstellung zu Stooping nach dem Gespräch ab. Der ausführliche Interview-Leitfaden befindet sich im Anhang B. Die Fragen des Leitfadens wurden in einer flexiblen Reihenfolge gestellt und eventuell ausgelassen (falls eine Person bereits ausführlich genug antwortete). Das Probeinterview wurde mit einem ersten Entwurf des Interview-Leitfadens durchgeführt. Nach dem Probeinterview mussten allerdings keine großen Anpassungen am Leitfaden vorgenommen werden, so dass das Interview in die Stichprobe mit aufgenommen wurde. Aufgrund von begrenzten finanziellen Mitteln wurde die Stichprobe für die Interviews nach dem Konzept des „Convenience Sampling[s]" ausgewählt (Cassell, 2015). Für die Studie wurden also Personen über persönliche Kontakte ausgewählt. Die Stichprobe wurde jedoch so gewählt, dass alle befragten Personen Stooping bereits einmal durchgeführt haben oder es in Erwägung ziehen würden. Diese Vorauswahl wurde getroffen, da die Stichprobe eine Grundgesamtheit von allen Personen widerspiegeln soll, die Stooping betreiben. Die Interviews fanden über zwei Wochen verteilt persönlich auf dem Universitätscampus oder bei den befragten Personen zuhause statt. Da teilweise auch Personen mit Wohnsitzen in anderen Städten oder Ländern befragt wurden, wurden auch Interviews über ein Online-Gespräch (WebEx) oder über ein Telefongespräch durchgeführt. Die

Gespräche hatten eine Länge von 25 bis 40 Minuten. Es wurden Personen aus mehreren deutschen Städten befragt (die meisten mit Wohnsitz in Berlin), außerdem auch zwei Personen mit Wohnsitzen in der Schweiz, in Österreich und in Israel. Die meisten Interviewpartner*innen sind 22 Jahre alt und sind Studierende. Genauere demografische Angaben befinden sich im Anhang A. Da die Interviewpartner*innen über persönliche Kontakte ausgewählt wurden, handelt es sich in Bezug auf das Alter und die Hauptberufstätigkeit um eine sehr homogene Gruppe.

3.1.3. Datenauswertungsmethode

Alle befragten Personen gaben zu Beginn der Interviews ihr mündliches Einverständnis, dass die Gespräche aufgezeichnet (Audio) und die anonymisierten Daten für Forschungszwecke verwendet werden. Nach der Aufnahme wurden die Audiodateien mit dem Datenauswertungsprogramm MAXQDA transkribiert. Bei der Transkription wurde das gesamte Gespräch niedergeschrieben. Lange Denkpausen wurden entsprechend kenntlich gemacht („*Ähm*“ oder „*()*“), Füllwörter oder Wortwiederholungen wurden jedoch nicht übernommen. Es wurde also ein Mittelweg zwischen dem von Oliver et al. (2005) beschriebenen Naturalismus (Beschreibung jeder Äußerung im Detail) und Denaturalismus (Entfernung von idiosynkratischen Elementen der Sprache, wie Stottern, Pausen oder Nonverbale) beim Transkribieren gewählt. Die ausführlichen Transkripte befinden sich im Anhang C. Nach der Transkription wurden die Gespräche ebenfalls mit dem Programm MAXQDA codiert. Beim Codieren wurde nach dem Konzept der thematischen Analyse vorgegangen. Hierbei werden die Textdaten mit einer Code-Vorlage organisiert, es können sich aber auch neue Codes im Laufe der Analyse ergeben und in die Vorlage aufgenommen werden (Cassell, 2015). Für diese konzeptgesteuerte Codierung (Gibbs, 2018) wurde ein Code-Handbuch als Vorlage aus der erarbeiteten Theorie und den Propositionen erstellt. Anhand dieses Handbuches wurden die Textdaten analysiert. Während der Analyse ergaben sich weitere Codes, so dass in einem iterativen Prozess auch diese Codes dem Handbuch hinzugefügt wurden. Eine Übersicht der erstellten Codes mit dazugehörigen Erklärungen befindet sich im Anhang D. Um die Qualität der Forschungsergebnisse sicherzustellen, wurde ein Augenmerk auf fünf von sechs Gütekriterien qualitativer Forschung nach Mayring (2016) gelegt. Diese Gütekriterien und wie sie bei der vorliegenden Untersuchung eingesetzt wurden sind dem Anhang E zu entnehmen.

3.2. Darstellung der Untersuchungsergebnisse

Im Folgenden werden die verschiedenen Antworten der Interviewpartner*innen zusammengefasst dargestellt. In einigen Fällen werden Auszüge aus den Interviews präsentiert, um aufgezeigte Verhaltensweisen besser verständlich zu machen.

3.2.1. Allgemeines Konsumverhalten

Die meisten befragten Personen geben an, dass sie insgesamt wenig Produkte konsumieren, da sie nur Produkte besitzen wollen, für die sie auch einen Zweck oder Nutzen haben. Dieser Zweck oder Nutzen wurde dementsprechend von allen Befragten als wichtige Eigenschaft eines zu erwerbenden Produktes angegeben. Weitere häufig genannte wichtige Produkteigenschaften sind ein guter Preis, ein ausgewogenes Preis-Leistungs-Verhältnis, eine gute Qualität, ein spezielles Design und Nachhaltigkeit. Neben der Angabe wenig zu konsumieren, berichten drei Personen, dass sie ausgeglichen konsumieren. Lediglich eine Person gibt an, viel zu konsumieren. Bei Dringlichkeit eines Produkts konsumieren viele Interviewpartner*innen neu, häufig wird hier der Grund der mangelnden Verfügbarkeit gebrauchter Produkte genannt. Sechs von acht Personen würden jedoch zuerst schauen, ob es das benötigte Produkt auch gebraucht gibt. Bezüglich der Entsorgung spenden alle Personen ihre noch brauchbaren Produkte. Außerdem werden aussortierte Produkte an Familie oder Freunde weitergegeben oder über Second-Hand Kanäle verkauft. In zwei Fällen werden die Produkte aufgewertet, so dass sie weiterhin Verwendung im eigenen Haushalt finden. Drei Personen erwähnten, dass sie ihre Produkte entsorgen, indem sie sie so an die Straße stellen, dass sie von anderen Personen durch Stooping erworben werden können.

3.2.2. Second-Hand Konsumverhalten

Second-Hand Kanäle, die die Interviewpartner*innen zum Erwerb gebrauchter Produkte nutzen, sind am häufigsten Online-Plattformen wie eBay oder Vinted. Darauf folgen Second-Hand Geschäfte, darunter Sozialkaufhäuser, bei denen die Erlöse für soziale Zwecke gespendet werden. Die meisten Personen geben ebenfalls an, schon einmal gebrauchte Produkte über den Second-Hand Kanal Flohmarkt erworben zu haben. Häufig wurde das Weitergeben von Produkten (vor allem Kleidungsstücke) an Familie und Freunde in der Kindheit genannt. Drei Personen erwerben auch heute noch gebrauchte Produkte durch Weitergabe aus ihrem sozialen Umfeld. Interessant ist, dass lediglich vier der Interviewten angeben, Stooping als Second-Hand Kanal zu nutzen, obwohl der Erwerb durch Stooping ein Auswahlkriterium der Interviewpartner*innen darstellte. Die befragten Personen wurden mit der Frage „Hast du schon einmal Produkte durch Mitnahme von der Straße erworben?“ ausgewählt. Dass Stooping jedoch nicht direkt als Kanal zum Erwerb gebrauchter Produkte genannt wird, könnte daran liegen, dass der Erwerb durch Stooping keine Regelmäßigkeit der befragten Personen darstellt. Neben einem Bedarf an Produkten (wie Dringlichkeit oder spezielle Anlässe), nennen vier Personen Langeweile oder das simple *„Gucken, was es gibt“ (Tim)* als Anlässe, Produkte über Second-Hand Kanäle zu konsumieren. Drei Personen besuchen Second-Hand Kanäle, wenn sie eine bestimmte Produktkategorie im Kopf haben, und nach diesen Produkten gebraucht suchen.

Obwohl nicht explizit danach gefragt, wurden einige Motivationen genannt, gebrauchte Produkte zu konsumieren. Am häufigsten nannten die Interviewten eine ökonomische Motivation, also dass gebrauchte Produkte günstiger zu erwerben sind. Eng verbunden mit der ökonomischen Motivation ist bei zwei Personen die Motivation gegen Verschwen-

dung, also dass Produkte nicht unnötigerweise weggeschmissen werden und noch einmal verwendet werden können. Weitere Motivationen der befragten Personen sind Nachhaltigkeit gebrauchter Produkte und Originalität. Rebecca bevorzugt gebrauchte gegenüber neuer Kleidung, da diese ihrer Meinung nach gesünder ist, weil sie keine Chemikalien aus der Herstellung mehr enthält. Eine weitere Motivation der Personen Produkte Second-Hand zu erwerben, stellt das Finden oder Entdecken von besonders schönen oder gesuchten Produkten dar. Klara empfindet *„Spaß, den man hat, wenn man Sachen raussucht“*.

Gebrauchte Produkte werden anders wahrgenommen als neue Produkte, so haben sie für drei Personen einen Charakter oder eine Seele. Mit ihnen wird häufig die Geschichte verbunden, wie man sie erworben hat oder wo sie herkommen. Diese Geschichte wird von Klara auch gerne an ihr soziales Umfeld weitergegeben und sie gibt an, gerne *„Sachen zu haben, die kein anderer hat“*. Außerdem empfinden Lukas und Melanie eine Verbundenheit zum*zur vorherigen Besitzer*in des Produkts.

Die Hälfte der Befragten gibt an, dass sie aufgrund von mangelnder Verfügbarkeit oder fehlender Möglichkeiten des Erwerbs, nicht dazu neigen, gebrauchte Produkte zu konsumieren. Als weitere Abneigung gebrauchten Produkten gegenüber wird von drei Personen die Unkenntnis über den*die vorherige*n Besitzer*in und die Furcht vor Verschmutzung oder Kontamination angegeben. Nur wenige Interviewpartner*innen stört die mangelnde Qualität oder die steigenden Preise von Second-Hand Produkten.

3.2.3. Stooping

Nachdem das Konsumverhalten gebrauchter Produkte im Allgemeinen dargestellt wurde, wird im Folgendem das Stooping-Verhalten der Interviewten näher beleuchtet. Bei der Durchführung der Interviews wurden Fragen verwendet, bei denen Stooping als ‚Mitnehmen von der Straße‘ bezeichnet wurde. Nur zwei der acht Interviewpartner*innen gaben am Ende des Gesprächs an, den Begriff ‚Stooping‘ zu kennen. Außerdem empfinden zwei Personen, dass das Phänomen des Mitnehmens von der Straße in ihrem Wohnort (Berlin) noch nicht weit verbreitet ist beziehungsweise in anderen deutschen Städten gängiger sei. Die Mehrheit der Interviewpartner*innen nimmt spontan etwas von der Straße mit, wenn sie zufällig daran vorbeilaufen und informieren sich nicht vorher auf Plattformen in den sozialen Medien, wo Produkte zur Verfügung stehen. Wenn sie auf ein Produkt auf der Straße stoßen, geben fünf Personen an, das Produkt nur mitzunehmen, wenn sie einen Bedarf oder Zweck dafür haben. Im Zusammenhang mit diesen Angaben steht, dass der Zweck oder Nutzen eines Produkts für die meisten Interviewpartner*innen eine wichtige Produkteigenschaft darstellt. So beschreibt Klara, dass sie nur Produkte von der Straße nimmt, für die sie selbst einen Nutzen hat, damit sie keinen Personen die Gegenstände *„[wegnimmt]“*, die sie vielleicht nötiger haben.

Gegenstände, die von den befragten Personen von der Straße mitgenommen werden, sind vor allem Möbel und Bücher, aber auch Kleidung, Pflanzen, Deko oder Haushaltsartikel. Eine Produktkategorie, die vier Personen explizit nicht von der Straße mitnehmen würden, ist Kleidung. Als Grund hierfür nennt z. B. Rebecca die Unkenntnis über den*die vorherige*n Besitzer*in und die Herkunft der Kleidung. Gefundene Produkte, die für eine Weile in den Haushalt integriert wurden, sind laut der Interviewpartner*innen nach einer Weile mindestens genauso viel wert wie neue Produkte. Dass die Produkte an der Straße zu verschenken oder zum Mitnehmen gedacht sind, wird allen Personen durch Kennzeichnungen an den entsprechenden Produkten klar. Die meisten Personen beschreiben auch eine Art Eindruck, den die Produkte vermitteln, da sie einzeln auf der Straße stehen. Darüber hinaus würden fünf Personen bei Gelegenheit nachfragen, ob die Produkte zu verschenken seien. Verena aus der Schweiz, Sophie aus Oldenburg und Klara aus Tel-Aviv beschreiben, dass es in ihrer Gegend gängig sei, Sachen an die Straße zu stellen beziehungsweise sie mitzunehmen.

Als häufigster Grund, Produkte durch Stooping zu erwerben, wurde der Gratiserwerb der Produkte genannt. Alle Personen nennen diese ökonomische Motivation als Grund, eine Sache mitzunehmen oder als Tatsache, über die sie sich besonders freuen. Als bestes gefundenes Produkt geben einige Personen Produkte an, deren Wert sehr hoch ist, für die sie aber nichts bezahlen mussten.

> *„[Ich fühle] Freude. Fast schon mehr Freude, als würde man etwas kaufen, weil man es gratis irgendwie bekommen hat.“ (Rebecca)*

Danach folgte die Bewahrung der Produkte vor einer Verschwendung. Produkte sind *„einen Versuch wert“ (Tim),* sollen *„eine zweite Chance [bekommen]“ (Verena), „gerettet werden“ (Klara)* oder sind *„zum Verrotten zu schade“ (Melanie)*. Bei einigen Personen spielt außerdem die Nachhaltigkeit in Bezug auf Ressourcenschonung bei der Rettung der Produkte eine wichtige Rolle.

> *„Ich finde das einfach schön, wenn man Sachen, die man selber nicht mehr benötigt, dem Kreislauf wieder zufügt. [. . .] Das hat ja alles viel Arbeit gemacht und ich fände es schade [. . .], wenn man einfach sowas auf den Müll schmeißt.“ (Melanie)*

Wie bereits erwähnt, nehmen fünf Personen Produkte von der Straße nur mit, wenn sie einen Zweck oder Nutzen dafür haben. Darüber hinaus empfinden einige Interviewte eine besondere Freude, weil sie die Produkte, die sie auf der Straße entdeckt haben, gerade benötigten. Sophie sagt:

> *„Wir hatten [...] vorher nach genau dieser Pflanze [. . .] immer Ausschau gehalten und waren auch in Baumärkten und in Pflanzengeschäften, um uns andere Pflanzen zu kaufen und da hat es uns auf jeden Fall gefreut [, dass wir sie gefunden haben].“*

Ein weiterer Grund, warum die Hälfte der befragten Personen Produkte von der Straße mitnimmt und der ihr Freude bereitet, ist die positive Überraschung. Wie bereits erwähnt, werden Produkte beim Stooping meist spontan erworben und nicht erwartet. Spaß wird dadurch empfunden, *„dass man [. . .] etwas sieht, das einem gefällt und dass man mitnehmen darf und es ist völlig so in dein Leben geplatzt“ (Verena)*. Mit dieser Motivation ist außerdem ein Glücksgefühl verbunden, das die Interviewpartner*innen empfinden, wenn sie ein Produkt auf der Straße entdecken. Dieses Gefühl kommt auf, da ein gewisser Zufall damit verbunden ist, dass die Personen das Produkt auf der Straße überhaupt erwerben können und es z. B. nicht schon von einer anderen Person mitgenommen wurde. Im Vergleich zum herkömmlichen Konsum neuer Produkte und dem Konsum gebrauchter Produkte über klassische Second-Hand Kanäle, empfindet die Hälfte der Befragten den Prozess der Konsumentscheidung als einfach. Sie beschreiben, dass *„eine komplizierte Entscheidungsfindung im Laden“ (Antonia)* durch eine simple Mitnahmeentscheidung ersetzt wird. Produkte werden außerdem von der Straße mitgenommen, um sie zu reparieren oder aufzuwerten. So kann ihnen *„ein zweites Leben geschenkt werden“ (Verena)* oder sie können zum *„Unikat“ (Tim)* gemacht werden. Drei Personen geben ebenso an, dass sie Spaß beim *„Durchstöbern der Kisten [auf der Straße]“ (Sophie)* haben und es sich wie eine Art *„Schatzsuche“ (Klara)* anfühlt. Weniger häufig genannte Motive sind die Qualität der Produkte, die auf der Straße zu finden sind (*„das Produkt [. . .] hat schonmal einen Haushalt durchgestanden und ist nicht komplett kaputt gegangen“ (Lukas)*) und ethische Gründe, also dass z. B. kein Fast Fashion beim Erwerb unterstützt wurde.

Ähnlich wie beim Konsum gebrauchter Produkte im Allgemeinen, werden mit Produkten, die durch Stooping erworben wurden, Geschichten verbunden. Die Mehrheit der Personen verbindet eine spezielle Erinnerung mit den Produkten. Erinnerungen können der (Gratis-) Erwerb oder die Geschichte des Produkts beziehungsweise der Vorbesitzer*innen sein. Diese Geschichten und Erinnerungen werden von den Interviewpartner*innen gerne ihren Freund*innen oder Mitbewohner*innen erzählt. Sie empfinden Stolz, auf gefunden Produkte von der Straße angesprochen zu werden. Die Befragten besitzen gerne Produkte, die niemand anderes hat, und verbinden damit eine positive Bewertung. Die Besonderheit und Originalität der durch Stooping erworbenen Produkten ist für die Befragten von hohem Wert.

> *„Und dazu kommt dann halt, dass man hoffentlich was hat, was gerade nicht jeder andere auch hat. Also, dass vielleicht der Blumentopf, den ich da eingesammelt hab, irgendein Erbstück von irgendeiner Oma oder so [ist] und ich [. . .] kann die Geschichte dazu erzählen [. . .].“ (Klara)*

Sieben von acht Personen empfinden eine Art Dankbarkeit den Personen gegenüber, die (brauchbare) Produkte an die Straße stellen. Lukas spricht sogar von einer „*Verbundenheit*“ und betont, dass sowohl die Person, die die Produkte herausstellt, als auch die Person, die sie mitnimmt, „*auf der gleichen Augenhöhe [sind]*“. Melanie, die selbst Produkte an die Straße stellt, berichtet von einem Gefühl der Gegenseitigkeit.

> *„[Ich] freue mich, wenn sich andere Leute freuen, und freue mich, wenn ich auch was Schönes finde, was ich gebrauchen könnte.“*

Im Laufe der Interviews hat sich herausgestellt, dass alle Personen mindestens einen Grund haben, Produkte von der Straße nicht mitzunehmen. Da sich diese Arbeit jedoch auf die Motivationen und (positiven) Resultate von Stooping konzentriert, werden die Gründe im Folgenden nur kurz thematisiert. Einige Personen berichten aufgrund der Spontanität des Erwerbs von logistischen Problemen, Produkte von der Straße mit nach Hause zu nehmen. Außerdem werden Produkte mitgenommen, für die Personen keinen Zweck oder Nutzen haben, und deren Erwerb sie später bereuen. Ein weiterer Aspekt, der stört, ist die Unbrauchbarkeit der Produkte auf der Straße. Entweder werden Produkte vor die Tür gestellt, die von vornherein nicht mehr brauchbar sind oder sie werden durch Wettereinflüsse (wie Regen) unbrauchbar. Die befragten Personen beklagen, dass so die Straßen mit entsorgtem Sperrmüll vollgestellt werden.

> *„Wo hört Sperrmüll auf, wo fängt zu verschenken an?“ (Lukas)*

Ebenso fürchten sich die Interviewpartner*innen vor einer Fehlinterpretation, also dass sie Produkte mitnehmen, die nicht zum Erwerb gedacht sind (z. B. bei einem Umzug oder einer Beantragung auf Sperrmüll-Abholung). Das Mitnehmen von Sperrmüll sei auch verboten. Im Vergleich zu klassischen Second-Hand Kanälen wird eine stärkere mangelnde Verfügbarkeit von eventuell gesuchten Produkten genannt und eine fehlende Kenntnis über den*die Vorbesitzer*in beziehungsweise eine fehlende Kontrolle der Qualität der Produkte. Zuletzt empfinden mehrere befragte Personen Scham, wenn sie dabei beobachtet werden, wie sie Produkte von der Straße mitnehmen.

4. Diskussion der Untersuchungsergebnisse

Nach der Darstellung der Untersuchungsergebnisse folgt nun die Diskussion dieser. Initial beschäftigte sich die qualitative Analyse damit, die erarbeiteten Propositionen und das aufgestellte Modell zu überprüfen, und weitere Motivationen und Resultate der Anti-Konsum Praxis Stooping zu ergründen.

4.1. Stooping als Anti-Konsum Praxis

Die ersten Kapitel dieser Arbeit beschäftigten sich damit, Stooping als Anti-Konsum Praxis und Second-Hand Konsum einzuordnen. Dass befragte Personen Stooping und Second-Hand Konsum *„in eine Kategorie gesteckt habe[n]“ (Rebecca)*, sich nach der Befragung aber der Unterschiede bewusst waren, spricht dafür, dass Stooping eine Art von Second-Hand

Konsum ist. Allerdings lässt sich diskutieren, ob Stooping wirklich als Anti-Konsum Praxis eingeordnet werden kann. Die Interviewpartner*innen begeben sich nicht immer gezielt auf eine Produktsuche über Stooping, weil sie Gründe besitzen, die gegen den herkömmlichen Konsum neuer Produkte sprechen. Vielmehr erwerben sie die Produkte nur spontan, wenn sie darauf stoßen. Ein möglicher Grund hierfür kann sein, dass der Erwerb durch Stooping in den Wohnorten einiger Interviewpartner*innen noch nicht sehr verbreitet ist und auch organisatorische Plattformen wie Instagram-Kanäle noch nicht bekannt oder etabliert sind. So ist es schwer, gezielt nach bestimmten Produkten zu suchen. Auch entsorgen die meisten Personen ihre Produkte, indem sie sie verkaufen oder spenden und nicht, indem sie sie an die Straße stellen, so dass sie durch Stooping zu erwerben sind. Die Stooping-Praktiken der Interviewten wären also nur als Anti-Konsum einzuordnen, wenn sie bewusst und absichtlich ausgeführt worden wären (Chatzidakis & Lee, 2013; Cherrier et al., 2011; García-de-Frutos et al., 2018). In weiteren Untersuchungen müsste erfragt werden, ob die Personen mit ihrem Erwerb über Stooping z. B. ein Zeichen gegen den Konsum neuer Produkte setzen wollten, z.B. gegen Fast Fashion oder die Verschwendung von Ressourcen (wie bereits in Interviews erwähnt). Ein Beispiel, welches nicht als Anti-Konsum Praxis einzuordnen wäre, wäre der Erwerb einer Musikbox durch Stooping von Lukas. Er hat den Neukauf des Produkts nicht in Betracht gezogen, da er keinen Bedarf für die Musikbox hatte. Trotzdem erwarb er das Produkt durch Stooping, jedoch nur weil es gratis war und er Lust hatte, es aufzuwerten.

4.2. Motivationen von Stooping

Da, wie bereits erwähnt, der Erwerb über Stooping nur spontan und nicht geplant stattfindet, beschränken sich die Motivationen für Stooping in den durchgeführten Gesprächen lediglich auf die Mitnahmeentscheidung der Produkte und nicht auf die Entscheidung, Stooping als Kanal für den Erwerb von Produkten zu wählen. Aus diesem Grund werden Produkte von der Straße nicht unbedingt neuen Produkten aufgrund der fehlenden monetären Gegenleistung vorgezogen. Vielmehr entscheiden sich die Interviewpartner*innen spontan, das Produkt von der Straße mitzunehmen, weil es gratis ist. Die befragten Personen geben jedoch an, sich mehr über Gratisprodukte zu freuen als über dieselben Produkte, die sie in einem Geschäft erworben hätten. Außerdem berichten sie von außergewöhnlichen Produkten, die sie durch Stooping günstig erwerben konnten, die sonst aber zu viel gekostet hätten (*„Retrostühle“ (Rebecca), „Pflanze“ (Sophie)*). Klara nennt zudem explizit die *„Schnäppchen“*, die sie durch Stooping machen kann. Da hauptsächlich Personen befragt wurden, die sich im Studium befinden und noch kein volles Einkommen erhalten, scheint der Preis eine wichtige Produkteigenschaft zu sein. Geld zu sparen und Schnäppchen zu machen (ökonomische Motive) stellen also eine Motivation der Interviewpartner*innen dar, Produkte durch Stooping zu erwerben beziehungsweise die Produkte spontan mitzunehmen. Proposition 1 kann also bestätigt werden. Diese Spontanität, die durch das ökonomische Motiv ausgelöst werden kann, kann jedoch auch unerwünscht sein. Die befragten Personen erwerben Produkte, für die sie eigentlich keinen Zweck haben und bereuen den Erwerb im Nachhinein.

Bezüglich der hedonistischen Motivation verhält es sich bei den Interviewten so, dass sie sich nicht aufgrund des Spaßes oder der Freude beim Erwerb auf eine gezielte Suche nach Produkten, die an der Straße stehen, begeben. Sie nehmen jedoch Produkte spontan mit nach Hause, weil sie im Moment der Entdeckung positiv überrascht sind und Glück verspüren, das (außergewöhnliche) Produkt gefunden zu haben. Sie sind überrascht, da sie im Voraus nicht wissen können, was und wann sie etwas auf der Straße finden werden. Drei Personen berichten außerdem, dass sie Spaß beim *„Durchstöbern der Kisten [auf der Straße]“ (Sophie)* haben und es sich wie eine Art *„Schatzsuche“ (Klara)* anfühlt, wenn sie spontan auf eine Kiste auf der Straße stoßen. Es macht also noch einmal einen Unterschied, ob die Produkte einzeln auf der Straße entdeckt werden, oder mehrere Produkte in Kisten liegen, und diese erst einmal durchgeguckt werden müssen. Bei allen drei Fällen stellen die Gefühle der Freude (positive Überraschung, Glück oder Schatzsuche) jedoch hedonistische Werte dar und sind eine Motivation, die gefundenen Produkte zu erwerben. Somit wurde auch Proposition 2 bestätigt.

Die befragten Personen begeben sich also nicht auf die ‚Jagd‘ (Charbonneau, 2008) und damit nicht auf eine gezielte Schatzsuche. Vielmehr berichten die Interviewpartner*innen von einer Einfachheit, Produkte durch Stooping zu erwerben *„ohne, dass du das jagen musst“ (Verena)*. Die Recherche nach speziellen Produkten in einem herkömmlichen Second-Hand Kanal oder für neue Produkte wird durch eine simple Mitnahmeentscheidung der Personen ersetzt. Diese Einfachheit ist für die befragten Personen ein Motiv, Produkte durch Stooping zu erwerben und stellt neben den erarbeiteten Propositionen eine weitere Motivation dar, Stooping zu betreiben. Auf der anderen Seite führt diese vereinfachte Konsumentscheidung zusammen mit den ökonomischen Motiven zum vermehrten unerwünschten spontanen Erwerb der Befragten.

Die Freude, die von den Personen beim Stooping empfunden wird, ist außerdem größer, wenn ein Produkt auf der Straße entdeckt wurde, für das auch ein Bedarf besteht. Ein Grund hierfür könnte sein, dass die Personen angeben, nur wenig Produkte zu konsumieren und nur solche, für die sie auch einen spezifischen Zweck oder Nutzen haben. Diesen Aspekt geben viele Befragte als Grund an, ein Produkt von der Straße mitzunehmen. Verbunden mit dieser Freude, ist die positive Überraschung und das Glück, das empfunden wird, wenn ein Produkt gefunden wird, für das auch ein Bedarf besteht.

> *„Wenn man das Glück hat, sich im Kopf zu denken, [...] ich möchte mir jetzt einen neuen Stuhl kaufen und dann findet man einen Stuhl, [...] hat man nochmal ein ganz anderes Gefühl, als wenn man [ihn] einfach nur so sieht.“ (Lukas)*

Da der Bedarf für ein Produkt jedoch eine notwendige Bedingung für den Erwerb ist, wird er hier nicht als Motivation für den Erwerb von Produkten durch Stooping berücksichtigt.

Nach dem ökonomischen Motiv am zweithäufigsten genannt ist das Motiv gegen die Verschwendung. Die Produkte können vor dem Müll *„gerettet werden“ (Klara)*, indem sie wiederverwendet, aufgewertet oder repariert werden. Somit wird die unnötige Verbreitung weiterer Produkte verhindert. Außerdem werden durch Stooping weniger natürliche Ressourcen verbraucht. Bezüglich der Verschwendung nannten zwar nicht alle befragten Personen explizit das Motiv der Nachhaltigkeit, dennoch kann die Bewahrung vor Verschwendung als Nachhaltigkeitsaspekt eingeordnet werden. Personen werden also auch durch Nachhaltigkeitsaspekte dazu motiviert, Produkte von der Straße mitzunehmen. Stooping stellt somit eine umweltorientierte Anti-Konsum Praxis dar, richtet sich also durch das Wiederverwenden von Produkten (Black & Cherrier, 2010) gegen den klassischen Konsum, um die Umwelt zu schützen (García-de-Frutos et al., 2018). Allerdings muss, wie bereits in Kapitel 4.1 erwähnt, überprüft werden, ob Stooping als umweltorientierter Anti-Konsum bewusst und absichtlich ausgeführt wurde (Chatzidakis & Lee, 2013; Cherrier et al., 2011; García-de-Frutos et al., 2018). Verbunden mit den Nachhaltigkeitsaspekten ist außerdem der Fakt, dass Personen Nachhaltigkeit als wichtige Produkteigenschaft angeben. Sie können bei durch Stooping erworbenen Produkten sichergehen, dass sie im Einklang mit ihren nachhaltig orientierten Prinzipien handeln. Neben den Nachhaltigkeitsaspekten wird außerdem von einer Person ein ethisches Motiv beschrieben, Produkte von der Straße mitzunehmen. Klara beschreibt, dass sie sich besser fühlt, Produkte durch Stooping zu erwerben oder in anderen Second-Hand Kanälen zu kaufen, da sie so ein Zeichen gegen Fast Fashion setzen kann. Dieses Motiv wurde zwar nur von einer Person genannt, ist aber eng verwandt mit dem Motiv der Nachhaltigkeit und der Bewahrung vor Verschwendung. Nachhaltigkeitsaspekte und ethische Motive stellen also Motivationen dar, Produkte durch Stooping zu erwerben. Proposition 3 wird bestätigt und durch ethische Motive ergänzt.

Eine weitere Motivation, die sich durch die Interviews ergeben hat, ist das Aufwerten oder Reparieren von Produkten, die an der Straße gefunden wurden. Die Interviewpartner*innen möchten die Produkte aufwerten, da sie kaputt sind oder weil sie *„Kreativität walten [. . .] lassen [möchten]“ (Verena)*. Etwas selbst herzustellen oder selbst aufzuwerten (DIY), kann eine Aktivität sein, aus der Personen Kraft, Sinn und Motivation schöpfen (Vosse, 2013). Die Produkte werden durch eine Aufwertung besonders. Diese Besonderheit und Originalität können zu einer verstärkten Verbundenheit zum Produkt oder einer positiveren soziale Wahrnehmung führen, worauf noch genauer in Kapitel 4.3 eingegangen wird.

Lediglich eine Person nannte die Qualität gebrauchter Produkte als ein Motiv, Produkte von der Straße mitzunehmen. Produkte haben eine gute Qualität, da sie zwar schon einmal verwendet wurden, jedoch nicht bereits kaputt gegangen sind. Interessanterweise wird diese Produkteigenschaft auch einmal (von einer anderen Person) als Motiv genannt, Second-Hand Shopping im Allgemeinen zu betreiben. Dem gegenüber stehen Personen, die die mangelnde Qualität, und dadurch Unbrauchbarkeit gebrauchter Produkte, auch von Produkten auf der Straße im speziellen, kritisieren. Dieser Widerspruch zeigt, dass die Motivationen, Produkte zu erwerben individuell und subjektiv sind. Die Qualität gebrauchter Produkte wird dennoch neben den erarbeiteten Propositionen als weitere Motivation aufgenommen, Produkte durch Stooping zu erwerben.

4.3. Resultate von Stooping

Nachdem nun die Motivationen erläutert wurden, widmet sich dieses Kapitel den Resultaten und den damit verbundenen Werten des Erwerbs durch Stooping und dem Besitz durch Stooping erworbener Produkte. Einige Personen berichten von einer Art Gemeinschaft, die sich um das Verschenken und Mitnehmen von Produkten dreht. So erzählt Rebecca von einer *„Kultur“* in ihrem Hausflur, und auch Verena, Sophie und Klara berichten, dass es in ihrer Wohngegend gängig sei, Sachen an die Straße zu stellen und sie mitzunehmen. Die Interviewpartner*innen nehmen also an gemeinsamen Aktivitäten mit anderen Personen teil, berichten aber nicht explizit von einem verstärkten Gefühl der Gemeinschaft, das sie motiviert, an den Aktivitäten teilzunehmen oder das ein Resultat dieser ist. Sie berichten eher von einer Art Verbundenheit den Personen gegenüber, die brauchbare Produkte so an die Straße stellen, dass sie mitgenommen werden können. Sie schätzen es wert, dieselbe Einstellung wie andere Personen zu haben (z. B. in Bezug auf die Bewahrung der Produkte vor Verschwendung) und berichten von einem Gefühl der Gegenseitigkeit. Wenn sie eine Person dabei beobachten, wie sie etwas an die Straße stellt, können sie leicht mit ihr ins Gespräch kommen. Durch den Erwerb in der Nachbarschaft und nicht in einem Geschäft, beruht die Beziehung zwischen der Person, die die Produkte bereitstellt und der Person, die die Produkte mitnimmt, nicht auf Geld, sondern vielmehr auf einer Gegenseitigkeit und Dankbarkeit. Ein Grund, weshalb die Personen von keinem Gefühl der Gemeinschaft sprechen, könnte sein, dass Stooping zwar als kollaborativer Konsum eingeordnet werden kann, der gemeinsame Konsum jedoch zu unterschiedlichen Zeitpunkten stattfindet (Felson & Spaeth, 1978). Durch die mangelnde Präsenz anderer Personen, die Stooping betreiben, kann die Gemeinschaft nicht richtig wahrgenommen werden. Durch die vermehrte Nutzung von Organisationsseiten, wie Instagram-Kanälen, könnte sich ein stärkeres Gefühl der Gemeinschaft entwickeln, da die Aktionen anderer Personen sichtbarer würden. Proposition 4 kann also nicht direkt bestätigt werden, sondern muss dahin gehend geändert werden, dass Personen, die Stooping betreiben, eine verstärkte Verbundenheit zwischen Verkäufer*in und Käufer*in und nicht primär ein verstärktes Gefühl der Gemeinschaft empfinden.

Neben dem Gefühl der Verbundenheit zwischen Verkäufer*in und Käufer*in, verbinden die Interviewpartner*innen

Erinnerungen mit ihren durch Stooping erworbenen Produkten. Diese Erinnerungen können die Erwerbssituation (Gratiserwerb), das Aufwerten des Produkts, die Geschichte des Produkts oder der Vorbesitzer*innen sein. Die Erinnerungen machen das Produkt besonders und es besitzt für die befragten Personen einen höheren Wert. Sie fühlen sich mit ihnen mehr verbunden als mit neuen Produkten.

> *„[Ich verbinde] definitiv immer noch Freude und halt eine andere Erinnerung [mit dem Produkt], als wenn ich jetzt in den nächsten Möbelladen [...] gehe oder online mir das bestelle. Auf jeden Fall [steckt da] schon eine andere Geschichte und eine individuellere Tat oder Gedanke dahinter."*
> *(Sophie)*

Allerdings verbinden die Befragten nicht direkt ihre Beziehungen zu anderen Personen oder ihre eigene Identität mit den Produkten, so wie von Kleine et al. (1995) beschrieben. Sie arbeiten die Produkte von der Straße jedoch auf, wodurch Energie in sie investiert wird. Diese Energie kann dazu führen, dass Objekte noch stärker als Erweiterung des Selbst angesehen werden (Belk, 1988) und damit eine stärkere Verbundenheit zu ihnen existiert (Kleine et al., 1995). Tim z. B. mag es, wenn Produkte eine Geschichte haben und er durch die Bearbeitung eines Produkts dieses zum *„Unikat"* machen kann und es dann *„persönlich"* wird. Proposition 5 wird somit größtenteils bestätigt.

Diese mit dem Produkt verbundenen Geschichten werden gerne von den befragten Personen an ihr soziales Umfeld weitergegeben. So werden Informationen über das Produkt weitergegeben und wahrgenommen. Vor allem außergewöhnliche und originelle Produkte werden gerne den Mitbewohner*innen oder Freund*innen gezeigt. Durch die spezielle Erwerbsgeschichte können sich die Befragten abheben und mit ihrem sozialen Umfeld interagieren, indem sie die Geschichte zu dem Produkt erzählen. Sie verbinden mit der Besonderheit der Produkte eine positive soziale Bewertung. Hier überschneiden sich die Ergebnisse aus den Befragungen mit der Literatur von Bellezza (2022) und Richins (1994). Nach der Frage, ob sie versuchen, ein spezielles Erscheinungsbild abzugeben, geben jedoch einige Personen an, dass ihnen dieses nicht wichtig sei und wenn, nur für sich selbst. Außerdem gehen sie nicht explizit darauf ein, Produkte wegen einer positiveren sozialen Wahrnehmung zu erwerben. Ihre Konsumentscheidung wird also nicht bewusst von ihrem sozialen Umfeld beeinflusst. Doch obwohl sie nicht offen zugeben, dass ihnen eine positive soziale Wahrnehmung wichtig ist, deuten die Angaben, dass sie solche Geschichten gerne teilen, darauf hin, dass sie daran interessiert sind, wie sie von anderen wahrgenommen werden.

Ein weiterer Widerspruch der Befragten zeigt sich darin, dass sie gerne Geschichten über ihre außergewöhnlichen Produkte teilen, jedoch unangenehm berührt sind, wenn sie beim Stooping dieser Produkte von anderen Personen beobachtet werden. Sie befürchten eine Vorverurteilung der Menschen, die sie nicht genau kennen. Dagegen teilen sie mit Stolz, wo sie das Produkt erworben haben, wenn sie von ihrem sozialen Umfeld darauf angesprochen werden. Dies könnte daran liegen, dass sie sich sicher sein können, dass ihr soziales Umfeld ihre Situation nicht missversteht und sie ihren Erwerb über Stooping besser rechtfertigen können. Verbunden mit dem Resultat der positiven sozialen Wahrnehmung ist außerdem die Motivation, entdeckte Objekte aufzuwerten. Die auf der Straße gefundenen Produkte werden mitgenommen, um sie später aufzuwerten und dadurch noch einzigartiger und außergewöhnlicher zu machen (Reiley & DeLong, 2011). Diese Originalität der Produkte führt wiederum zu einer verstärkten positiven sozialen Wahrnehmung, da nicht nur die Geschichte über den Erwerb des Produkts erzählt werden kann, sondern auch die Geschichte über das Aufwerten des Produkts. Auch Proposition 6 kann also teilweise bestätigt werden, der genaue Zusammenhang zwischen Stooping und der Wahrnehmung des sozialen Umfelds müsste allerdings noch weiter untersucht werden.

Abbildung 2 stellt eine Übersicht der von den befragten Personen genannten Motivationen und Resultate von Stooping dar. Einige zuvor vermutete Motivationen und Resultate wurden ersetzt beziehungsweise angepasst. Außerdem haben sich durch die Befragungen auch weitere Motivationen von Stooping (siehe Kapitel 4.2) ergeben.

4.4. Limitationen und künftiger Forschungsbedarf

Der Fokus dieser Arbeit lag darauf, die Motivationen und Resultate von Stooping zu ergründen. Mögliche Motivationen und Resultate wurden im Voraus mithilfe verschiedener Literatur, insbesondere Literatur zu Second-Hand-Konsum, erarbeitet. Auch die Interviewpartner*innen wurden während des Gesprächs gefragt, wie sich ihr Second-Hand Konsumverhalten gestaltet. Es wurden zudem einige Motivationen und Resultate von Second-Hand Konsum genannt, obwohl die Befragung nicht explizit darauf abzielte. Diese Motivationen und Resultate sind in Kapitel 3.2.2 nachzulesen. Jedoch sprengt es den Umfang dieser Arbeit, auf einen genauen Vergleich zwischen den Motivationen und Resultaten von Stooping und Second-Hand Konsum einzugehen. In dieser Richtung besteht also noch weiterer Forschungsbedarf.

Zudem sind die Motivationen und Resultate von Stooping oft miteinander verknüpft und häufig individuell und subjektiv. Das zeigen Widersprüche, die sich bei den Befragungen ergeben haben. So nannten z. B. einige befragte Personen die mangelnde Qualität gebrauchter Produkte als Kritikpunkt, während andere Befragte die Qualität von Second-Hand Produkten als Motiv erwähnten, diese zu erwerben. Diese Zusammenhänge müssten noch weiter vertieft und untersucht werden. Auch der Zusammenhang zwischen Stooping und der Wahrnehmung des sozialen Umfelds müsste noch genauer untersucht werden. Außerdem ergaben sich während der Interviews einige Gründe, die gegen Stooping sprechen. Diese wurden in Kapitel 3.2.3 bereits dargestellt, jedoch noch nicht genauer diskutiert.

In dieser Arbeit fokussieren sich die erarbeiteten Motivationen, Produkte über Stooping zu erwerben, lediglich auf die Mitnahmeentscheidung der Produkte und nicht auf die

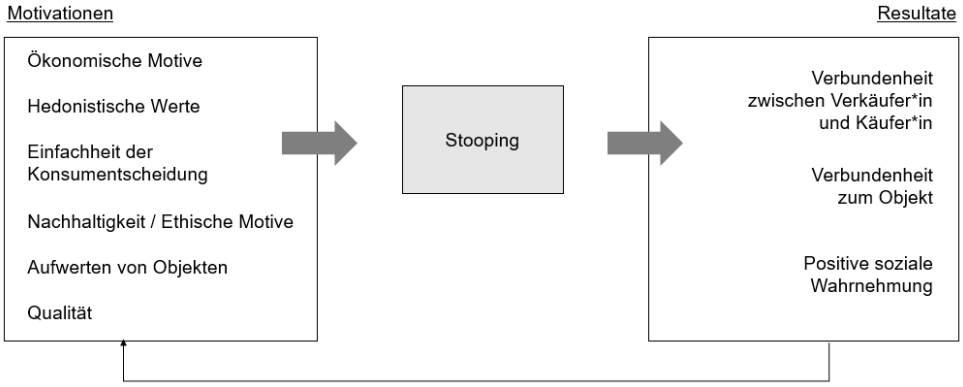


Abbildung 2: Übersicht der Motivationen und Resultate von Stooping nach Auswertung der Untersuchungsergebnisse (eigene Darstellung)

grundsätzliche Entscheidung, Stooping als Kanal für den Erwerb von Produkten zu nutzen. Die Interviewpartner*innen berichteten, dass der Konsum von Produkten spontan stattfindet, und der Erwerb über den Kanal ‚Stooping' von den Befragten nicht geplant bzw. im Voraus überdacht wurde. Da das Phänomen Stooping wie bereits erwähnt unter Konsument*innen noch nicht weit verbreitet ist, ist Stooping als Kanal für die nachhaltige Beschaffung von Produkten vermutlich noch nicht sehr präsent. Sobald dies jedoch der Fall ist, würde sich auch eine wissenschaftliche Betrachtung der Entscheidung, Stooping als Kanal für den Erwerb von Produkten zu nutzen, anbieten.

Außerdem könnte ein weiteres Augenmerk daraufgelegt werden, ob die Motivationen und Resultate von Stooping für verschiedene Produktkategorien unterschiedlich sind (z. B. Kleidung im Vergleich zu Möbelstücken). Es müsste auch untersucht werden, ob Stooping als Anti-Konsum Praxis eingeordnet werden kann, der Konsum über Stooping also absichtlich und bewusst ausgeführt wurde (Chatzidakis & Lee, 2013; Cherrier et al., 2011; García-de-Frutos et al., 2018). Für die Befragungen kamen viele Studierende in Frage, da diese oft Stooping betreiben. Eine Erklärung hierfür wäre, dass sie wenig Geld besitzen und ökonomische Motive haben, Stooping zu betreiben. Personen, die ihr Einkommen freiwillig reduzieren, vereinfachen ihr Leben stärker als Personen, die ihren Konsum einschränken (Etzioni, 1998). So sind Studierende mit wenig Einkommen eher bereit, Produkte von der Straße mitzunehmen, als Personen mit höherem Einkommen, die sich die Produkte aus Bequemlichkeit neu kaufen können. Hier stellt sich die Frage, ob dieser Konsum dann noch als Anti-Konsum Praxis einzuordnen wäre, da er grundsätzlich keinen freiwilligen Verzicht auf neue Produkte darstellt. Es besteht nämlich ein Unterschied zwischen dem ‚Nicht-Kaufen-Können' und dem ‚Nicht-Kaufen-Wollen' von Produkten (Close & Zinkhan, 2009). Durch das Convenience Sampling wurden nicht nur überwiegend Studierende befragt, sondern auch Personen aus persönlichen Kontakten. Eine quantitative Untersuchung könnte einen möglichen Bias Effekt bereinigen. Durch eine Ausweitung der Stichprobengröße könnten außerdem eventuell weitere Motivationen und Resultate entdeckt werden, die mit einer quantitativen Studie bestätigt werden könnten. Bezüglich des Konsumprozesses über Stooping wäre es weiterhin interessant, den Fokus vom Erwerb und der Nutzung gebrauchter Produkte von der Straße auf die Entsorgung auszuweiten. Es könnte untersucht werden, was die Motivationen und Resultate vom Stellen von Produkten an die Straße sind.

5. Fazit

Stooping ist eine Form des alternativen Konsums (und Second-Hand Konsums), die unter bestimmten Bedingungen auch als eine Praxis des Anti-Konsums einzuordnen ist. Zwar ist der Begriff noch neu und unbekannt, das Phänomen des An-die-Straße-Stellens von Produkten und das Mitnehmen dieser jedoch in einigen Städten bereits geläufig. Durch eine qualitative Untersuchung in Form von Interviews wurde in dieser Arbeit herausgefunden, dass Personen durch ökonomische Motive, hedonistische Werte, der Einfachheit der Konsumentscheidung, der Nachhaltigkeit (beziehungsweise ethische Motive), das Aufwerten von Produkten und der Qualität der Produkte motiviert werden, Stooping zu betreiben. Die Resultate des Konsums gebrauchter Produkte über Stooping können eine Verbundenheit zwischen Verkäufer*in und Käufer*in, eine Verbundenheit zum Objekt und eine positive soziale Wahrnehmung des sozialen Umfelds sein.

Durch diese Erkenntnisse wird klarer, wie sich das Konsumverhalten rund um Stooping gestaltet. Stooping hat sich bereits als alternative soziale Praxis zum herkömmlichen Konsum neuer Produkte etabliert. Damit es jedoch als vollwertige soziale Innovation angesehen werden kann, muss Stooping von der breiten Bevölkerung langfristig angenommen werden (Jaeger-Erben et al., 2015). Es ist bisher schwierig, Produkte durch Stooping zu erwerben, da organisationale Kapazitäten fehlen, über die sich potenzielle Konsumierende informieren können. Bis jetzt gibt es nur wenige kleinere Plattformen für diesen Zweck in den sozialen Medien. Um Stooping zu einer echten Alternative zum Konsum neuer Produkte zu machen, könnten Unternehmen weitere (Online-) Plattformen für die Organisation von Stooping bereitstellen und davon profitieren. Durch eine weitreichende Vernetzung

der Personen, die Produkte an die Straße stellen, sie entdecken und sie konsumieren, kann Stooping für Konsumierende zugänglicher gemacht werden. So könnte der Erwerb von Produkten über Stooping in Zukunft nicht mehr nur spontan, sondern auch geplant und bewusster stattfinden. Des Weiteren können Unternehmen durch den Verkauf von Produkten für die Aufwertung von Objekten profitieren. Mehrere Interviewpartner*innen werten ihre gefundenen Objekte von der Straße auf und benötigen hierfür Hilfsmittel. Der Verkauf von DIY-Produkten ist mit einer breiten Palette an ergänzenden Produkten und Dienstleistungen verknüpft (Hornik & Feldman, 1982). Zudem können politische Maßnahmen (Jaeger-Erben et al., 2015), wie die Entkriminalisierung beziehungsweise Legalisierung von Stooping helfen, diese Praxis für mehr Personen erreichbarer zu machen. Mithilfe all dieser Maßnahmen können mehr Produkte durch Stooping erworben und die Verschmutzung der Straßen durch Müll und unbrauchbare Produkte verhindert werden. So kann der Konsum von Produkten über Stooping nicht mehr nur eine Erweiterung der Erwerbsmöglichkeiten von Produkten sein, sondern als vollwertige, nachhaltige und verantwortungsvolle Alternative zum herkömmlichen Konsum neuer Produkte gelten.

Literatur

Albinsson, P. A., Wolf, M., & Kopf, D. A. (2010). Anti-consumption in East Germany: Consumer resistance to hyperconsumption. *Journal of Consumer Behaviour*, *9*(6), 412–425. https://doi.org/10.1002/cb.333

Albinsson, P. A., & Yasanthi Perera, B. (2012). Alternative marketplaces in the 21st century: Building community through sharing events. *Journal of Consumer Behaviour*, *11*(4), 303–315. https://doi.org/10.1002/cb.1389

Balderjahn, I., Seegebarth, B., & Lee, M. S. W. (2021). Less is more! The rationale behind the decision-making style of voluntary simplifiers. *Journal of Cleaner Production*, *284*, 124802. https://doi.org/10.1016/j.jclepro.2020.124802

Bardhi, F., & Arnould, E. J. (2005). Thrift shopping: Combining utilitarian thrift and hedonic treat benefits. *Journal of Consumer Behaviour*, *4*(4), 223–233. https://doi.org/10.1002/cb.12

Belk, R. (1988). Possessions and the extended self. *Journal of Consumer Research*, *15*(2), 139–168. https://doi.org/10.1086/209154

Belk, R. (2010). Sharing. *Journal of Consumer Research*, *36*(5), 715–734. https://doi.org/10.1086/612649

Bellezza, S. (2022). Distance and alternative signals of status: A unifying framework. *Journal of Consumer Research*. https://doi.org/10.1093/jcr/ucac049

Black, I. R., & Cherrier, H. (2010). Anti-consumption as part of living a sustainable lifestyle: Daily practices, contextual motivations and subjective values. *Journal of Consumer Behaviour*, *9*(6), 437–453. https://doi.org/10.1002/cb.337

Cassell, C. (2015). Conducting Research Interviews for Business and Management Students. https://doi.org/10.4135/9781529716726

Charbonneau, J. S. (2008). *Social responsibility and women's acquisition of secondhand clothing* [Philosophie]. Colorado State University. https://mountainscholar.org/handle/10217/3202

Chatzidakis, A., & Lee, M. S. W. (2013). Anti-consumption as the study of reasons against. *Journal of Macromarketing*, *33*(3), 190–203. https://doi.org/10.1177/0276146712462892

Cherrier, H., Black, I. R., & Lee, M. (2011). Intentional non-consumption for sustainability: Consumer resistance and/or anti-consumption? *European Journal of Marketing*, *45*(11/12), 1757–1767. https://doi.org/10.1108/03090561111167397

Cherrier, H., Szuba, M., & Özçağlar-Toulouse, N. (2012). Barriers to downward carbon emission: Exploring sustainable consumption in the face of the glass floor. *Journal of Marketing Management*, *28*(3/4), 397–419. https://doi.org/10.1080/0267257X.2012.658835

Close, A. G., & Zinkhan, G. M. (2009). Market-resistance and Valentine's Day events. *Journal of Business Research*, *62*(2), 200–207. https://doi.org/10.1016/j.jbusres.2008.01.027

De Young, R. (1996). Some psychological aspects of reduced consumption behavior: The role of intrinsic satisfaction and competence motivation. *Environment and Behavior*, *28*(3), 358–409. https://doi.org/10.1177/0013916596283005

DeepL SE. (n. d.). DeepL. Verfügbar 1. Februar 2023 unter https://www.deepl.com/translator

Etzioni, A. (1998). Voluntary simplicity: Characterization, select psychological implications, and societal consequences. *Journal of Economic Psychology*, *19*(5), 619–643. https://doi.org/10.1007/978-3-662-03900-7_1

Felson, M., & Spaeth, J. L. (1978). Community structure and collaborative consumption: A routine activity approach. *American Behavioral Scientist*, *21*(4), 614–624. https://doi.org/10.1177/000276427802100411

García-de-Frutos, N., Ortega-Egea, J. M., & Martínez-del-Río, J. (2018). Anti-consumption for environmental sustainability: Conceptualization, review, and multilevel research directions. *Journal of Business Ethics*, *148*, 411–435. https://doi.org/10.1007/s10551-016-3023-z

Gibbs, G. R. (2018). Thematic coding and categorizing (2. Aufl.). https://doi.org/10.4135/9781526441867

Gould, S. J., Houston, F. S., & Mundt, J. (1997). Failing to try to consume: A reversal of the usual consumer research perspective. *Advances in Consumer Research*, *24*, 211–216. https://web.s.ebscohost.com/ehost/pdfviewer/pdfviewer?vid=0&sid=7797653f-cef7-4b7c-ad3f-abf652cc26b6%40redis

Guiot, D., & Roux, D. (2010). A second-hand shoppers' motivation scale: Antecedents, consequences, and implications for retailers. *Journal of Retailing*, *86*(4), 355–371. https://doi.org/10.1016/j.jretai.2010.08.002

Hornik, J., & Feldman, L. P. (1982). Retailing implications of the do-it-yourself consumer movement. *Journal of Retailing*, *58*(2), 44–63. https://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=4666901&site=ehost-live

Jaeger-Erben, M., Rückert-John, J., & Schäfer, M. (2015). Sustainable consumption through social innovation: A typology of innovations for sustainable consumption practices. *Journal of Cleaner Production*, *108*, 784–798. https://doi.org/10.1016/j.jclepro.2015.07.042

Kleine, S. S., Kleine, R. E., & Allen, C. T. (1995). How is a possession „me"or „not me"? Characterizing types and an antecedent of material possession attachment. *Journal of Consumer Research*, *22*(3), 327–343. https://doi.org/10.1086/209454

Lambert, M. (2004). 'Cast-off wearing apparell': The consumption and distribution of second-hand clothing in northern England during the long eighteenth century. *Textile History*, *35*(1), 1–26. https://doi.org/10.1179/004049604225015620

Lee, M., Roux, D., Cherrier, H., & Cova, B. (2011). Anti-consumption and consumer resistance: Concepts, concerns, conflicts, and convergence. *European Journal of Marketing*, *45*(11/12), 1680–1687. https://doi.org/10.1108/ejm.2011.00745kaa.001

Luchs, M., Naylor, R. W., Rose, R. L., Catlin, J. R., Gau, R., Kapitan, S., Mish, J., Ozanne, L., Phipps, M., Simpson, B., Subrahmanyan, S., & Weaver, T. (2011). Toward a sustainable marketplace: Expanding options and benefits for consumers. *Journal of Research for Consumers*, (19). http://jrconsumers.com/Academic_Articles/issue_19/Sustainability_academic4.pdf

Mayring, P. (2016). Gütekriterien qualitativer Forschung (6. Aufl.), 140–148.

McCracken, G. D. (1990). Culture and consumption: New approaches to the symbolic character of consumer goods and activities.

momox AG. (2022). Second hand fashion report 2022. https://momox.biz/images/Pressemitteilungen/2022/Booklet_Second_Hand_Fashion_Report_momox_fashion_2022.pdf

Oliver, D. G., Serovich, J. M., & Mason, T. L. (2005). Constraints and opportunities with interview transcription: Towards reflection in qualitative research. *Social Forces*, *84*(2), 1273–1289. https://doi.org/10.1353/sof.2006.0023

Pavlakoudis, R. (2022). Was tun Verbraucher für die Circular Economy? Verfügbar 30. Januar 2023 unter https://www.softwareadvice.de/blog/3113/was-verbraucher-fur-zirkulaere-wirtschaft-tun

Putnam, R. D. (2000). Bowling alone: The collapse and revival of American community.

Reiley, K., & DeLong, M. (2011). A consumer vision for sustainable fashion practice. *Fashion Practice*, *3*(1), 63–83. https://doi.org/10.2752/175693811X12925927157054

Richins, M. L. (1994). Valuing things: The public and private meanings of possessions. *Journal of Consumer Research*, *21*(3), 504–521. https://doi.org/10.1086/209414

Roulston, K. (2014). Analysing interviews (U. Flick, Hrsg.). https://doi.org/10.4135/9781446282243

Steffen, A. (2017). Second-hand consumption as a lifestyle choice. https://doi.org/10.15501/978-3-86336-918-7_16

stoopingberlin. (n. d.). Instagram-Fotos und -Videos. Instagram-Profil. https://www.instagram.com/stoopingberlin/

StoopingNYC. (n. d.). Instagram-Fotos und -Videos. Instagram-Profil. https://www.instagram.com/stoopingnyc/

Szmigin, I., Carrigan, M., & McEachern, M. G. (2009). Conscious consumer: Taking a flexible approach to ethical behaviour. *International Journal of Consumer Studies*, *33*(2), 224–231. https://doi.org/10.1111/j.1470-6431.2009.00750.x

TheSorryGirls. (2022). The new thrifting no one is talking about. Verfügbar 3. Januar 2023 unter https://youtu.be/z0sjAVMINfs

thredUP. (2020). 2020 Resale report. https://www.thredup.com/resale/2020/static/thredup-resaleReport2020-42b42834f03ef2296d83a44f85a3e2b3.pdf

thredUP. (2022). 2022 Resale report. https://www.thredup.com/resale/static/2022-resaleReport-full-92a77020598ceca50f43227326100cc2.pdf

Vosse, C. (2013). Do-it-yourself and sustainable consumption - The changing meaning of do-it-yourself. *Forschungsjournal Soziale Bewegungen*, *26*(1), 90–95. https://doi.org/10.1515/fjsb-2013-0112

Waight, E. (2013). Eco babies: Reducing a parent's ecological footprint with second-hand consumer goods. *International Journal of Green Economics*, *7*(2), 197–211. https://doi.org/10.1504/IJGE.2013.057444

Weil, C. (1999). Secondhand chic: Finding fabulous fashion at consignment, vintage, and thrift shops.

Williams, C. C., & Windebank, J. (2002). The 'excluded consumer': A neglected aspect of social exclusion? *Policy & Politics*, *30*(4), 501–513. https://doi.org/10.1332/030557302760590422

Wolf, M., & McQuitty, S. (2013). Circumventing traditional markets: An empirical study of the marketplace motivations and outcomes of consumers' do-it-yourself behaviors. *Journal of Marketing Theory and Practice*, *21*(2), 195–210. https://doi.org/10.2753/MTP1069-6679210205

Xu, Y., Chen, Y., Burman, R., & Zhao, H. (2014). Second-hand clothing consumption: A cross-cultural comparison between American and Chinese young consumers. *International Journal of Consumer Studies*, *38*(6), 670–677. https://doi.org/10.1111/ijcs.12139

Zahid, N. M., Khan, J., & Tao, M. (2022). Exploring mindful consumption, ego involvement, and social norms influencing second-hand clothing purchase. *Current Psychology*. https://doi.org/10.1007/s12144-021-02657-9

Zavestoski, S. (2002). Guest editorial: Anticonsumption attitudes. *Psychology & Marketing*, *19*(2), 121–126. https://doi.org/10.1002/mar.10005

Junior Management Science 9(3) (2024) 1681-1699

Junior Management Science

www.jums.academy
ISSN: 2942-1861

When Does Marketing & Sales Collaboration Affect the Perceived Lead Quality? – The Moderating Effects of IT Systems

Lukas Hilke

Ruhr University Bochum

Abstract

In the realm of corporate dynamics, lead management remains a relatively underexplored subject, despite its increasing significance and annual resource allocation. This study addresses the enigmatic "sales lead black hole" by investigating the influence of enhanced collaboration between marketing and sales on the perceived quality of marketing-generated leads. A research model was crafted to delve into this relationship and assess the impact of contemporary IT systems on collaboration, subsequently bolstering the perceived lead quality. Findings reveal that active collaboration in planning lead management activities and exchanging information elevates the acceptance of marketing-generated leads, prompting increased follow-up engagement by sales personnel. IT systems play a pivotal role in fostering such collaboration, amplifying its effect on the perceived quality of leads. This research contributes vital insights for scholars by dissecting key drivers of perceived lead quality and proposing solutions for the sales lead black hole. For practitioners, the study offers actionable implications to enhance subjective perceptions of marketing-generated leads, curbing resource wastage through improved follow-up strategies.

Keywords: it-systems; lead management; marketing & sales collaboration; perceived lead quality; sales engagement

1. Introduction

Digitalization has greatly changed the way companies search for information and interact with each other (Järvinen & Taiminen, 2016). Customers are better informed than ever before, with up to 60% of the typical B2B buying process already completed before they contact a manufacturer or sales representative (Adamson et al., 2012). Nevertheless, or maybe especially due to this fact, new customer acquisition remains one of the biggest challenges for marketing and sales (Torkornoo, 2020). To keep up with the changed customer behavior, companies are spending up to 10% of their revenues on marketing initiatives to position themselves as a viable supplier as early as possible in their customers' buying journey (Gartner, 2021).

Especially online marketing is an important and fast-growing trend. Not only because old-fashioned lead channels, like trade shows, need to be replaced due to the pandemic, but also because online leads are often far more cost-efficient than traditional lead channels (Fröhlich, 2021; Team Linchpin, 2022). Around 70% of marketing resources are spent on digital marketing initiatives which, combined with the right automation tools, can lead to a strongly increasing number of leads (Gartner, 2021; HubSpot, 2022a; Moran, 2022; Mrohs, 2021). Leads are often described as the lifeblood of companies, which is why an increasing number of leads can be seen as positive and is often used by marketing as a metric for measuring the success of their activities (Monat, 2011; Wenger, 2021).

However, this metric does not consider the important follow-up of these leads. In practice, up to 70% of leads generated by marketing are not followed up by sales. This phenomenon is also commonly described as the "sales lead black hole" (Michiels, 2009; Sabnis et al., 2013). Due to this insufficient follow-up, companies constantly loose ready-to-buy customers (Hasselwander, 2006; Sabnis et al., 2013). The poor follow-up can be attributed to a variety of causes like missing information, delayed processes or miscommunication and distrust between marketing and sales (van der Borgh et al., 2020).

DOI: https://doi.org/10.5282/jums/v9i3pp1681-1699

Many researchers have tried to develop mechanisms to prioritize leads and evaluate their quality (D'Haen & van den Poel, 2013; Monat, 2011; Yan et al., 2015). Unfortunately, there is no uniform consensus on the criteria that determines the quality of a lead. In practice, qualification is usually based on intuition, supposed competence, and heuristic rules (D'Haen & van den Poel, 2013; Jolson, 1988). Therefore, a salesperson's follow-up effort is largely determined by their perception of lead quality, or more precisely their perception of the marketing department's lead prequalification process (Sabnis et al., 2013).

The factors that can influence the perceived lead quality have not yet been discussed. One of the most important factors can be identified as the collaboration between marketing and sales. Miscommunication, unclear processes and definitions represent a major hurdle for companies in establishing efficient lead management processes (Michiels, 2009). In particular, if sales is not involved in the planning and processes are untransparent, this can have a massive impact on the perceived lead quality, as marketing and sales may have different expectations of leads (Malshe & Sohi, 2009b). The influence of technology is creating numerous opportunities which enable the collaboration between marketing and sales to be even more efficient (Järvinen & Taiminen, 2016; Wiersema, 2013). IT-systems play a central role in optimizing the flow of communication and information and create new opportunities for collaboration between the two departments (Speier & Venkatesh, 2002). Good collaboration between marketing and sales should be essential to ensure common quality standards in lead management and to avoid the waste of resources. Especially in light of the increasing amount of resources spent on online marketing to generate leads and new business, it should be closer examined how good collaboration could influence a salesperson's perception of marketing-generated leads and how IT-systems enable these effects.

This paper will contribute to the lead management literature in filling this research gap and showing that marketing and sales collaboration has an important impact on the lead management process. First, the perceived lead quality will be confirmed to have a significant impact on a salesperson's lead follow-up effort. Second, the degree of joined planning and information sharing will both be identified as predictors of the perceived lead quality. Third, a positive influence of IT-systems on the collaboration between marketing and sales, as well as a moderating influence from IT-systems on the relationship between joint planning and perceived lead quality will be identified.

Based on these findings, interesting implications can be derived for research and practice. The results shed light on the key drivers of the perceived quality of marketing-generated leads and provide a more accurate understanding of the causes of and solutions to the sales lead black hole. The study also helps managers better understand what affects their team's follow-up rate and how they can improve it.

The remainder of this paper proceeds as follows: First, the conceptual framework underlying the study is presented. Then, previous literature on lead management, marketing and sales collaboration, and IT-systems is reviewed. Based on this, hypotheses are developed and tested in an empirical study, which is presented in the following section. Finally, the research findings and conclusions are discussed, including implications for both theory and practice.

2. Conceptual Framework

The research framework is illustrated in Figure 1. The framework assumes a direct influence from the collaboration of marketing and sales, represented by joined planning and information sharing, on the perceived quality of marketing-generated leads. The perceived lead quality will then have a direct impact on the follow-up efforts of marketing-generated leads. Furthermore, IT-systems, represented by the quality of lead information in the systems and the systems support in prioritization and planning, are expected to increase the collaboration between marketing and sales and also to moderate their relationship with the perceived lead quality.

The following sections review the literature on lead management, marketing and sales collaboration, and IT-systems before developing hypotheses and empirically testing the models.

3. Literature Review

3.1. Lead Management

Lead management is a topic of increasing importance, which however has been mostly neglected in research (van der Borgh et al., 2020). The term 'lead' describes a potential customer who has expressed an interest in a company's products or services, regardless of whether this is an existing customer or a new customer (Monat, 2011). Therefore, lead management describes the process by which potential buyers are developed into customers and is part of the sales funnel (Cooper & Budd, 2007; D'Haen & van den Poel, 2013). The sales funnel categorizes potential customers based on their buying stage, illustrating the ongoing narrowing and selection from all potential customers interested in a company's products and services to those customers who actually make a purchase (Cooper & Budd, 2007; Järvinen & Taiminen, 2016). The exact form of the sales funnel, as well as the number and arrangement of the different phases, differs from study to study (Järvinen & Taiminen, 2016). A frequently used framework is the model from D'Haen and van den Poel (2013), which divides the sales funnel into four phases: *Suspects*, *Prospects*, *Leads,* and *Customers*. Taking into account cross- and up-selling opportunities with existing customers, the last phase can also be replaced with *Deals*, turning the classic funnel model into a loop into which existing customers can re-enter (Järvinen & Taiminen, 2016; Patterson, 2007).

The process starts with the generation of leads, which is usually conducted by the marketing department. After the

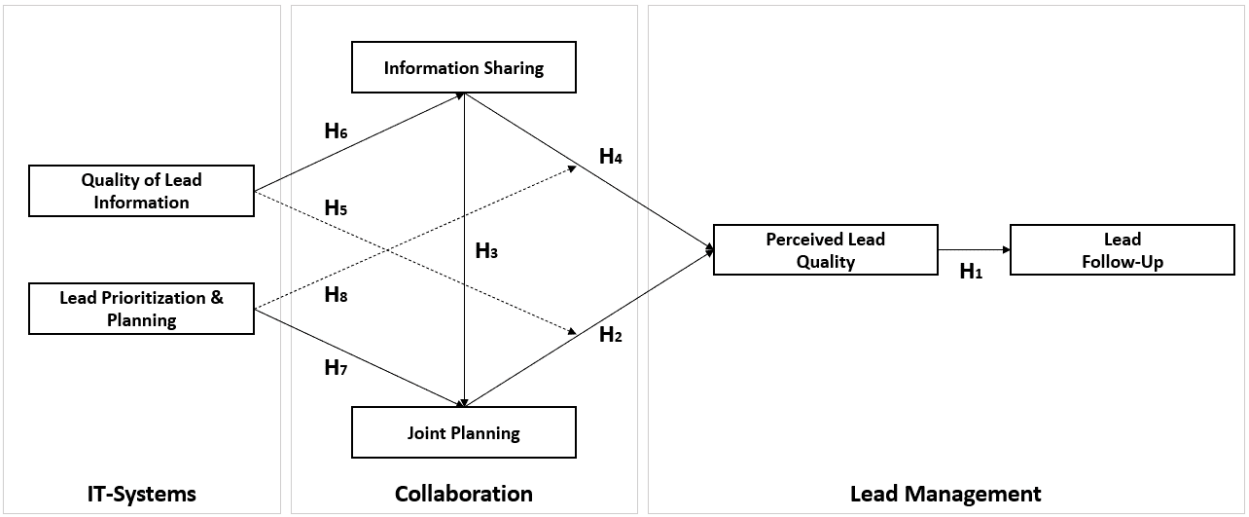


Figure 1: Illustration of the Conceptual Framework

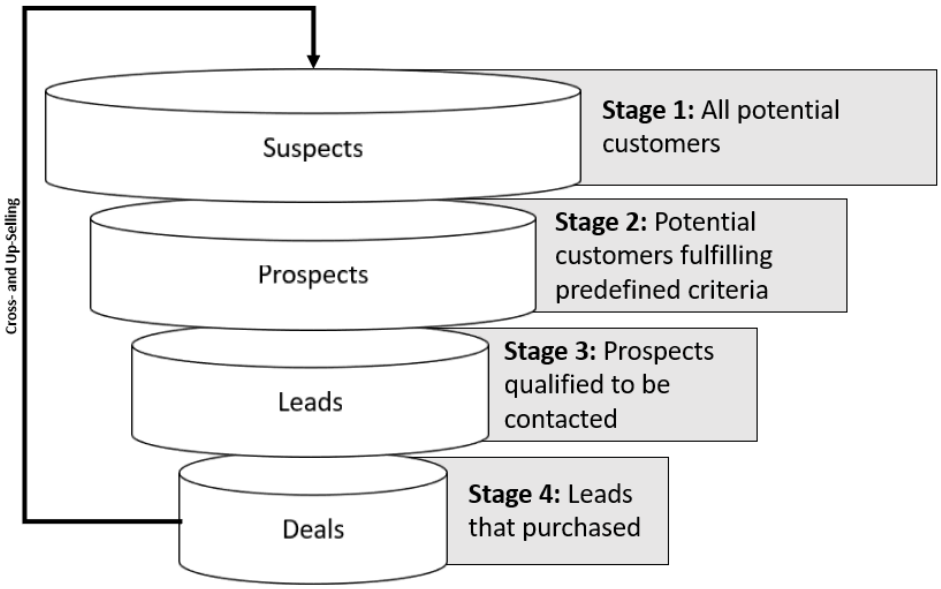


Figure 2: Sales funnel Framework (adapted from Järvinen and Taiminen (2016) and D'Haen and van den Poel (2013))

initial generation, leads are ideally pre-qualified and prioritized based on company-specific criteria before the leads are passed on to sales for follow-up (van der Borgh et al., 2020). Sales reps are normally expected to contact every lead they receive from the marketing department, but in reality, studies show that up to 70% of those leads are never contacted by sales (Michiels, 2009; Sabnis et al., 2013). Sales often argues that marketing-generated leads lack potential and that the quality is uncertain and unobservable (Banerjee & Bhardwaj, 2019; Oliva, 2006).

This perception can be traced down to a variety of different reasons. One problem is that leads often contain limited information about the prospect, making it difficult for marketers and salespeople to assess potential value and make informed decisions (Järvinen & Taiminen, 2016). Another common reason are inefficient and manual processes causing delays and preventing opportunities to smoothly transition from lead to sale (Michiels, 2009). Delays in the lead process can have a damaging effect on the chances of success in lead follow-up. The more time elapses after a customer inquiry, the less likely it is that a deal will be closed (Oldroyd et al., 2011; Smith et al., 2006).

These two reasons are likely connected to a third problem, the collaboration between marketing and sales. Various studies suggest that the gap between marketing and sales leads to a variety of problems in the lead management process (D'Haen et al., 2016; van der Borgh et al., 2020). Lack of coordination and unclear processes and definitions in the cooperation between sales and marketing represent a significant obstacle in lead management for companies (Michiels,

2009). In addition, insufficient involvement of sales in the planning of lead management activities can influence the perceived quality of the prequalification and, accordingly, the perceived lead quality (Malshe & Sohi, 2009b).

All those reasons can lead to a negative reputation of marketing-generated leads, resulting in reduced follow-up efforts and a waste of resources on both sides.

3.1.1. Lead Quality

To overcome the bad reputation of marketing-generated leads, researchers and practitioners have tried to develop models to evaluate the quality of leads and estimate the likelihood of a successful sale (D'Haen & van den Poel, 2013; Monat, 2011; Yan et al., 2015). The pre-qualification process is of particular importance in order to protect sales reps from a flood of poorly qualified and unpromising leads (Hise & Reid, 1994). When poorly qualified leads are handed over to sales, this has a strong negative impact on the sales funnel. First, sales reps waste their valuable time, in which they could have focused on selling products, following up on hopeless leads (Bradford et al., 2016; D'Haen & van den Poel, 2013). At the same time, the poor quality of leads reduces the motivation of salespeople to follow up on these leads, so that in the long run they are most likely to significantly reduce or even stop following up. (Sabnis et al., 2013).

Lead quality can be divided into objective and subjective quality. So far research and practice have mainly focused on the objective lead quality developing models to identify the most promising prospects (e.g., D'Haen and van den Poel, 2013; Michiels, 2009). The objective quality of a lead can be determined by various criteria; some of the most important include: the source of the inquiry (e.g., website, trade show, telemarketing), the need and urgency of the prospect, the budget and authority of the lead, the willingness to provide information, whether the contact was initiated by the company or the prospect, and whether the prospect has done business with the company before and/or fits the profile of a key account (Järvinen & Taiminen, 2016; Jolson, 1988; Jolson & Wotruba, 1992; Monat, 2011).

Unfortunately, the weighting and specification of the criteria varies from company to company and cannot be generalized (Monat, 2011). Furthermore, not all needed information about a lead is freely available, rather it is only known once an employee has contacted the prospect (Banerjee & Bhardwaj, 2019). As a result, companies are often forced to rely on publicly available information that is readily available but does not necessarily provide insight into the contact's interest in the company's products (Järvinen & Taiminen, 2016; Long et al., 2007). This is a major disadvantage, as signals of interest in a company's products are considered the most important indicators of purchase intentions among prospects (Bhattacharyya, 2014; Järvinen & Taiminen, 2016).

The objective quality of leads is therefore difficult to measure and cannot easily be observed by the sales reps. Consequently, salespeople are motivated by their subjective perception of the lead prequalification process and lead quality. van der Borgh et al. (2020), for example, found that both the speed and quality of lead assignment have an inverted U-shaped relationship with lead follow-up and must be consistent to achieve positive results. Their findings were based on the fact that if leads were frequently misassigned, and if the assignment of leads was too fast or too slow, salespeople would get the impression that the entire process, and thus the leads themselves, were of poor quality. If the quality of the prequalification process is low in salespeople's perception, they are more likely to focus on their self-generated, familiar leads than on leads from marketing (Sabnis et al., 2013). In their study, Sabnis et al. (2013) showed that better perceptions of the quality of the prequalification process led to higher expectations of success and greater willingness in following up on leads from marketing among salespeople.

In consequence, the focus should shift from the determinants of objective lead quality to the determinants of subjective lead quality instead, as this perception has a direct impact on the follow-up of leads from marketing. The exact factors influencing subjective lead quality have not yet been empirically investigated, but possible factors can be derived from related literature (Järvinen & Taiminen, 2016; Mero et al., 2020; Ohiomah et al., 2019; Wiersema, 2013).

3.2. Marketing & Sales Collaboration

Collaboration between marketing and sales seems to be one of the biggest problem drivers in lead management. In practice and in research, it has been repeatedly found that the cooperation between marketing and sales is not always harmonious and constructive (Biemans et al., 2010; Rouziès et al., 2005). Especially in lead management, sales reps often complain about the poor quality of marketing-generated leads, and marketing in turn complains about the poor follow-up efforts from the sales team (Biemans et al., 2010; Sabnis et al., 2013). The literature names a wide variety of reasons for the conflicts between marketing and sales, such as: different objectives (Strahle et al., 1996), poor communication and coordination in planning (Colletti & Chonko, 1997; Matthyssens & Johnston, 2006), different perspectives and thought worlds (Beverland et al., 2006; Homburg, Jensen, & Krohmer, 2008), lack of interfunctional integration (Rouziès et al., 2005), and lack of clarity about the roles and responsibilities of the other side (Biemans et al., 2010; Homburg & Jensen, 2007; Le Meunier-FitzHugh & Piercy, 2011; Malshe & Sohi, 2009b).

And yet, successful collaboration between marketing and sales can have a strong positive impact on the effectiveness of activities and the overall business performance of a company (Homburg & Jensen, 2007; Le Meunier-FitzHugh & Piercy, 2011; Rouziès et al., 2005). The interface between marketing and sales should actually be well-equipped for effective collaboration because marketing and sales both deal with potential customers. Marketing is tasked with supporting sales, running campaigns, and building a consistent brand image, and sales is responsible for more tactical tasks such as contacting customers, implementing the strategies, and closing deals (Biemans et al., 2010). There are even certain activities that can only be carried out effectively through coordinated

efforts of sales and marketing, among which lead management can be counted (Rouziès et al., 2005; Schmitz et al., 2020).

The literature provides various approaches to improve collaboration between marketing and sales in order to achieve this ideal state. Oliva (2006), for example, found that a common language as well as organizational and systematic links are of particular importance. A precise definition and understanding of core terms, such as leads, must be created among all team members. Especially companies in which the roles of marketing and sales and the entire sales process are clearly defined seem to achieve much better results. Malshe and Sohi (2009a) showed that if marketing makes strategic decisions independent of sales, sales reps tend to view these as irrelevant. Therefore, it is important to receive a buy-in from sales which can be achieved when marketing and sales engage in joint planning, operate in a fact-based environment, and share information (Malshe & Sohi, 2009a). Le Meunier-FitzHugh and Piercy (2011) noted that marketing and sales collaboration is positively associated with business performance and that information sharing, supportive attitudes, joint planning, and aligned goals are essential to receive results.

The communication of customer and market information across departments was identified as a key factor for an organization's responsiveness to customer needs because marketing and sales can provide each other with invaluable information to better address and target customers. (Colletti & Chonko, 1997; Kirca et al., 2005). This interfunctional communication was also identified by Hulland et al. (2012) as an important factor, but they also showed that increased communication is only beneficial if marketing and sales both perceive their interactions as fair. It is therefore not enough to increase the frequency of meetings and the exchange of information, instead the communication between marketing and sales must be perceived as collaborative and consultative (Dawes & Massey, 2005; Kahn & Mentzer, 1998).

Homburg, Jensen, and Krohmer (2008) found that a high level of knowledge sharing, structural linkage between departments, and a high level of expertise in both departments had a positive impact. In addition, Homburg, Jensen, and Krohmer (2008) showed that teams were more successful when sales adopted a more long-term perspective, matching marketing's naturally longer-term perspective. This was also confirmed by Homburg and Jensen (2007), who showed that although different perspectives can have a positive effect on market performance in some cases, the quality of cooperation between marketing and sales suffers as a result.

Lack of communication, coordination, and transparency in lead management can lead to doubts about the quality of lead pre-qualification and thus the quality of marketing-generated leads, as there has been no buy-in from sales into these processes (Malshe & Sohi, 2009a; Malshe et al., 2017) and marketing and sales may have different requirements and expectations of a lead (Malshe & Sohi, 2009b).

Improving collaboration through increased sharing of information and joint planning could therefore help to improve perceived lead quality and also resolve the problems of delayed processes and lack of information. Since marketing and sales can then better coordinate the processes and goals in lead management, and both sides will know exactly what lead information is needed and how delays in processes can be avoided.

Despite the importance of the topic and the well-researched marketing and sales interface, the impact of collaboration in lead management has not yet been empirically investigated and should therefore urgently be investigated in more detail.

3.3. IT-System Support

The potential of IT-systems in sales and marketing has been recognized early on (Collins, 1985). Today IT-systems like Lead Management (LMS), Salesforce Automation (SFA) or Marketing Automation Systems (MA) are already an integral part in the day-to-day activities of many companies, often fully integrated into one CRM-platform, offered by software vendors like HubSpot (HubSpot, 2022b). Organizations continuously invest in these state-of-the-art technologies to keep up with the competition by improving communication, information, and customer management (Jelinek et al., 2006; Peterson et al., 2011; Tanner et al., 2005). These technologies not only improve the quality and speed of information gathering (Speier & Venkatesh, 2002), but also help their users to become more efficient in managing customer interactions and information and to automate all kind of marketing and sales processes (Hunter & Perreault Jr., 2006; Zoltners et al., 2001). Vendors of different software solutions claim that the tools can help align companies marketing and sales interfaces, therefore improving and accelerating lead management processes (Järvinen & Taiminen, 2016). This would indicate that the support of IT-systems can improve the collaboration between marketing and sales and thus contribute to overcome the sales lead black hole.

In a recent survey 80% of the respondents noted that the implementation of marketing automation software increased their marketing and sales collaboration (Hannig et al., 2019), this might be due to the fact that such software requires constant interaction and sharing of information between both parties (Mrohs, 2021). By creating a 360°-view of the customer and their interactions, IT-systems provide the foundation for collaboration, knowledge creation and opportunity exploitation (Mero et al., 2020; Peterson et al., 2011; Plouffe et al., 2004). Marketing gets deeper insights into customer data and can better customize campaigns, whilst sales reps get more information about marketing-generated leads and their activities (Järvinen & Taiminen, 2016; Wiersema, 2013). This provision of more accurate and easily accessible information, combined with IT-support in selecting and qualifying leads, helps sales reps to better understand customer needs and customize proposals to cater these needs (Ahearne et al., 2007; Moutot & Bascoul, 2008; Román & Rodríguez, 2015).

Leads are information products and IT-systems help to organize, manage, and share this information effectively

(Ahearne et al., 2007; Kuruzovich, 2013). Nowadays sales and marketing have access to extensive amounts of data, but to be successful, they need to convert this data into useful information (Hunter & Perreault Jr., 2006). IT can support this process, for example, through the analysis of which lead sources are most profitable, and thus make the extensive amounts of data available useful for planning (Hunter & Perreault Jr., 2006; Kuruzovich, 2013; Tanner et al., 2005).

This means that both the collaboration of marketing and sales and IT-systems focus on improving the exchange of information and establishing clearly defined processes. The most important tasks of IT-systems in lead management are thus the provision and sharing of information, as well as transparent and, at best, automated processes (Ahearne et al., 2007; Ohiomah et al., 2019; Park et al., 2010). Since the introduction of IT-systems does not magically produce the desired status quo, complex IT-systems likely promote collaboration between marketing and sales to a degree that was not necessary before (Michiels, 2009). IT developments and shared systems between marketing and sales create the opportunity to reduce the perspective divide between both parties and allow them to better coordinate their activities, share information, and set shared goals (Hulland et al., 2012; Järvinen & Taiminen, 2016; Kotler et al., 2006; Tanner et al., 2005; Wiersema, 2013). Therefore, it could be concluded that IT-systems support the collaboration of marketing and sales in its core objectives and possibly enable it in its best possible form in the first place.

The influence of software on the lead management process has only been studied by Ohiomah et al. (2016, 2019). However, their study only considered a mediating relationship through an increased number of sales calls, which showed mixed results. Yet, the above discussion indicates that the impact of IT-systems on lead management is more likely to come from its impact on collaboration, which influences the perceived lead quality. Better IT-systems enable marketing and sales to more effectively scale lead processes, ensure timely and high-quality communication and make sure that existing data can be better used for planning and adjusting existing processes (Kuruzovich, 2013; Matthyssens & Johnston, 2006; Tanner et al., 2005). Therefore, it is likely that IT-systems influence the lead management process on the one hand by enabling better planning and information sharing between marketing and sales and on the other hand by enhancing the positive effects that collaboration has on perceived lead quality.

The research gap in the predictors of perceived lead quality is evident and needs to be examined in more detail (see Figure 3). The next section therefore investigates this research gap by developing a research model that examines the influence that marketing and sales collaboration has on the perceived quality of marketing-generated leads and how IT-systems enable better collaboration and affect its relationship with the perceived lead quality.

4. Hypotheses Development

Previous research has identified the perceived lead quality as one of the major predictors for sales rep's lead follow up (Sabnis et al., 2013). As the topic of lead management was neglected for a long time, it is not surprising that the factors that can influence this perceived lead quality have remained unexplored so far. This paper will address this research gap by investigating the collaboration of marketing and sales as one predictor of perceived lead quality. Additionally, the potential influences of modern-day IT-systems on collaboration and its relationship with the perceived lead quality will be investigated.

In order to analyze the effects of collaboration on perceived lead quality and identify possible moderating effects, it is first necessary to theorize how these influences may work. Therefore, it is useful to draw on empirical findings from previous research on lead management, collaboration, and IT-systems.

First, consistent with previous research, perceived lead quality is expected to influence the follow-up effort of sales reps. Then, we will explore how collaboration between marketing and sales may influence the perceived lead quality. Afterwards, we hypothesize about how different aspects of IT-systems can have an influence on collaboration and how they could moderate the relationship between collaboration and perceived lead quality.

4.1. Lead Follow-up Efforts

The follow-up of leads is an essential part of a salesperson's day-to-day activities (Ohiomah et al., 2019; Pullins et al., 2017). According to Sabnis et al. (2013) lead-follow-up characterizes the customer acquisition effort a salesperson dedicates to either self-generated or marketing-generated leads. It describes the ability of a salesperson to thoroughly follow up on leads and maintain contact with those leads until the completion of a sale or the abandonment of the lead (Ohiomah et al., 2016).

A sales rep's follow-up efforts are largely determined by his motivation, which can be divided into intrinsic and extrinsic motivation. Intrinsic motivation describes motivation that arises from the activity itself and from there emanates from the employee himself, while extrinsic motivation arises from the result of the activity as well as from external stimuli (Deci & Ryan, 2000; Pullins, 2001; Ryan & Deci, 2000).

One of those external stimuli has been identified as the perceived quality of the lead prequalification (Sabnis et al., 2013). Prequalification is the process of reviewing newly generated leads and deciding whether to pass them on to sales reps or further nurture them by the marketing department (Michiels, 2009; Mrohs, 2021; Sabnis et al., 2013). As described earlier, the objective quality of leads is difficult to measure and cannot easily be observed by the sales reps. Therefore, sales reps depend on their perception of lead quality or more precisely their perception of marketing's ability

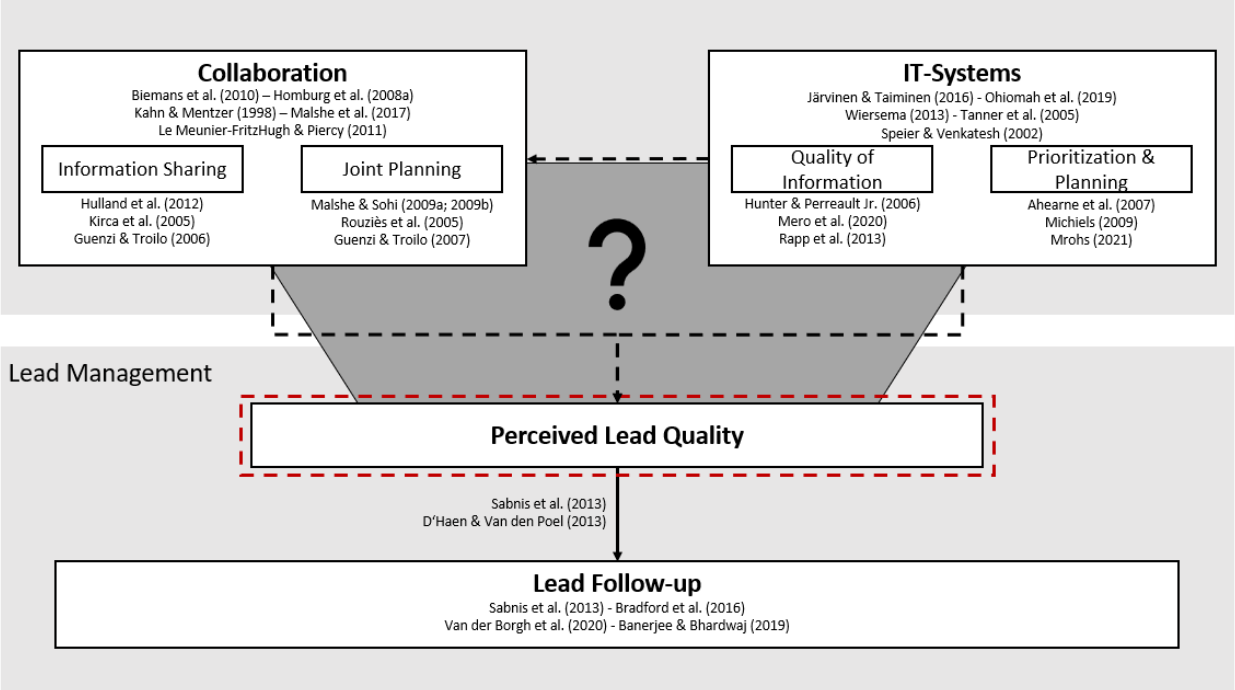


Figure 3: Illustration of the Research Gap

to screen and discard unattractive leads as well as to efficiently and timely assign leads to the right person (Sabnis et al., 2013; van der Borgh et al., 2020).

The assignment of unqualified leads has a negative effect on salespersons' motivation since they more often receive negative feedback, leading to disappointment and wasted resources (Jolson, 1988; Monat, 2011). If sales reps develop an unfavorable opinion of the prequalification process, they are more likely to spend their time on self-generated leads than on marketing-generated leads (Sabnis et al., 2013).

Consequently, an improvement in the perception of the prequalification process should increase the sales reps' expectations for success in pursuing these leads. If the sales rep becomes more confident that marketing has eliminated low potential prospects, this should in turn improve his extrinsic motivation to follow-up on those leads. Hence, in line with the results found by Sabnis et al. (2013), the following is hypothesized:

> *H1: A sales rep's perception of the quality of marketing-generated leads is positively associated with his follow-up efforts to these leads.*

4.2. Marketing & Sales Collaboration

Collaboration is defined as an affective and volitional process in which departments work together with mutual understanding, a shared vision, and shared resources to achieve common goals (Kahn & Mentzer, 1998). This mutual understanding, collective goals, and sharing of information and resources promotes goodwill between departments so that employees are more satisfied working with other departments (Kahn & Mentzer, 1998; Schrage, 1990; Souder, 1987).

As discussed above, literature already identified various ways how collaboration between marketing and sales can be improved. Almost all of them stress that joint planning and information sharing are of particular importance, which is why their influence will be investigated in more detail (Homburg, Jensen, & Krohmer, 2008; Le Meunier-FitzHugh & Piercy, 2011; Malshe & Sohi, 2009a). How these variables might affect the perception of lead quality will be discussed in the following sections.

4.2.1. Joint Planning

Joint planning between marketing and sales is defined as the co-development of goals, processes, and activities and in the case of this paper refers in particular to joint planning of lead management activities (Homburg, Jensen, & Krohmer, 2008). Existing scholars stressed that sales needs to be included in marketing strategy decisions and that marketing and sales must synchronize their strategic and tactical activities in order to design strategies that create, deliver & communicate superior customer value (Guenzi & Troilo, 2007; Malshe & Sohi, 2009b). This should be of particular importance in the case of lead management, where marketing typically is responsible for the first few steps, like creating brand awareness, marketing plans and leads for sales, and sales is then expected to execute the marketing plans and follow-up on those leads (Kotler et al., 2006).

In examining the development of marketing strategies, Malshe and Sohi (2009b) highlighted that for successful strategy development and implementation, both marketing and sales must be equally involved in the entire process. Similar results were found in another study by Malshe and Sohi

(2009a), confirming that sales needs to be involved in strategy creation and that marketing has to make sure to show them the bigger picture and the value that is added to their day-to-day activities. Mutual participation in the development of goals and processes is critical for sales acceptance of goals and their motivation to pursue them. It creates a sense of ownership and therefore makes it easier for marketing to receive sales buy-in (Malshe & Sohi, 2009b; Rouziès et al., 2005). Furthermore, through the joint development both sales and marketing should gain a better understanding and appreciation of the other functions issues and perspectives (Rouziès et al., 2005).

Joint planning of processes and goals should have an important impact on lead management, and especially on the lead qualification process. When marketing and sales determine criteria for qualified leads together as well as standardized processes for handling these leads, there should be fewer arguments about lead quality and follow-up practices (Järvinen & Taiminen, 2016). Sales reps therefore know what to expect from a lead, how to handle it, and who to turn to with feedback. Thus, one can hypothesize:

> *H2: Marketing and Sales engagement in jointly planning lead management activities is positively related to the perceived quality of marketing-generated leads.*

4.2.2. Information Sharing

Information Sharing encompasses the scope of cross-functional information dissemination and knowledge exchange (Homburg, Jensen, & Krohmer, 2008). The importance of information sharing has been highlighted in the literature on intraorganizational interfaces (e.g., Fisher et al., 1997) and especially in the literature on the sales and marketing interface (e.g., Homburg, Jensen, and Krohmer, 2008; Le Meunier-FitzHugh and Piercy, 2011).

Biemans et al. (2010), who studied different marketing and sales interface configurations, noted that in the most advanced interface configuration the integration of marketing and sales was realized through a mix of formal and informal communication, with people in both departments feeling motivated to exchange information. The sharing of lead and customer information between marketing and sales is important to stay responsive to changing customer needs (Hulland et al., 2012; Kirca et al., 2005). Both parties can provide each other with valuable information. For example, sales can provide marketing with information about which lead campaigns are generating the most promising prospects and suggest modifications that will fit customers' changing needs. Marketing, on the other hand, can help sales with information and tools that enable them to better target and approach customers (Colletti & Chonko, 1997).

The exchange of information creates a common understanding of situations and helps to bring sales and marketing on the same page (Biemans et al., 2010). Bidirectional communication, sharing of information, and feedback loops helps to identify problems in the lead process and allow both parties to fine-tune the existing strategy (Malshe & Sohi, 2009b; Wenger, 2021). From this, the following hypothesis can be concluded:

> *H3: The amount of information sharing between marketing and sales is positively associated with their engagement in jointly planning lead management activities.*

Malshe and Sohi (2009a) further noted that sharing information about feedback is of particular importance, especially when the feedback was not implemented, so that the other party knows what happened to their feedback and why it may not have been acted upon. This ensures that the dialogue and sharing of information between marketing and sales is maintained and does not break down. Fisher et al. (1997) even noted that setting norms for the sharing of information can significantly contribute to the creation of such an ongoing dialogue.

Better information sharing helps sales reps to become more efficient, since the provision of more accurate customer information enables salespeople to better customize proposals to the unique needs and concerns of their leads (Moutot & Bascoul, 2008; Ohiomah et al., 2019; Park et al., 2010).

Well-informed salespeople are perceived to have a higher level of commitment and trust in the existing structures (Matthyssens & Johnston, 2006; Siguaw et al., 1994). Guenzi and Troilo (2006) found that sharing of information fosters increased effectiveness and efficiency of market knowledge development and decision-making, while supporting an organizational climate of commitment and trust.

It can therefore be concluded that if marketing and sales frequently share information, it should not only promote joint planning behavior, but also create an environment in which sales reps have less doubts about the quality of the lead prequalification and of the leads themselves. This is because they receive more information about leads, which allows them to better approach those prospects. Additionally, they get a better understanding of what happens with their feedback, that it is heard and that strategies are adjusted if needed, which generates a feeling of trust and commitment in the existing processes, leading to the following hypothesis:

> *H4: The amount of information sharing between marketing and sales is positively related to the perceived quality of marketing-generated leads.*

4.3. IT-System Support

As described earlier, IT-systems simplify the daily tasks of marketing and sales while creating new opportunities for increased collaboration between the two parties. They can help bring marketing and sales together, as the systems can help them to better understand and trust each other's contribution (Tanner et al. 2015). The following describes how IT-system support, represented by the systems quality of lead information and the systems support in prioritization and planning, could improve collaboration, and further enhance the impact from collaboration on perceived lead quality.

4.3.1. Quality of Lead Information

The quality of lead information describes the usefulness, clarity, and accessibility of information about marketing-generated leads in a company's IT-systems. Nowadays, companies have access to a vast amount of data about their current and prospective customers and IT-systems can help to turn these data into useful information (Hunter & Perreault Jr., 2006). According to Ahearne et al. (2007) they enhance the richness, complexity, and mobility of information and knowledge by increasing the communication speed, information availability and remote accessibility.

Effective information is a meaningful input for successful planning of lead management activities. IT-systems can facilitate and enable increased information effectiveness, making it available not only for sales but also for planning activities (Hunter & Perreault Jr., 2006). When different IT-systems are synchronized, the data can provide a complete record of customer interactions in a timely and readily accessible manner, creating a holistic picture of customers and sales operations (Mero et al., 2020; Tanner et al., 2005). The ideal situation is a clear and easily accessible unified database that provides an 360° view of the customer. Marketing and sales are then provided with a better decision-making base in the joint creation of lead management activities, which also creates a higher degree of consistency in marketing and sales work (Mero et al., 2020; Mrohs, 2021).

The information gains through well-presented and high-quality data in those systems can help everyone to better understand the needs and purchasing abilities of the marketing-generated leads and how to best capture these opportunities (Moutot & Bascoul, 2008; Ohiomah et al., 2019; Park et al., 2010) also mentioned that one of the biggest benefits of IT-systems is that they help to learn more about customers and leads and can ultimately shape the way how those customers are approached. They found that IT usage was positively associated with market information processing, which indicates that IT-systems allow marketing and sales to work with information more quickly and effectively. Moreover, good and organized IT-systems also help to focus on the most important information, which enables everyone to develop winning strategies in less time (Rapp et al., 2008).

All this should create a better foundation to discuss and jointly plan strategies in lead management. A better quality of lead information in the systems should enable all parties to better coordinate and jointly adjust processes, as increased data quality and accessibility allows faster and more accurate decision-making. This should also generate a higher level of trust into the processes agreed upon, as they are backed by better information. Accordingly, the following hypothesis can be formulated:

> *H5: The positive effect of joint planning on perceived lead quality is expected to be higher the better the quality of lead information in the IT-systems is.*

Furthermore, when the IT-systems provide Marketing and Sales with more useful and qualitative data and give them easy and quick access to this improved data, it is likely that Marketing and Sales therefore begin to better understand the value of high-quality information. As a result, they should be more willing to share information and knowledge at the customer level to ensure a certain level of data quality in the systems. (Tanner et al., 2005).

Well organized IT-systems make sure no relevant data is lost or overlooked (Engle & Barnes, 2000; Matthyssens & Johnston, 2006). Additionally, those organized systems allow everyone to identify, gather and share relevant information more easily (Rapp et al., 2008). Especially when different sources are combined to one source of truth, relevant information no longer needs to be tediously gathered from different sources or queries, making it easier to find, handle, and analyze for everyone (Järvinen & Taiminen, 2016; Kotler et al., 2006; Wiersema, 2013). Accordingly, as the organization and quality of data in IT-systems becomes better the threshold for sharing relevant information between marketing and sales should decrease, as the information in the systems is more complete and easier to access and handle. Thus, it can be hypothesized that:

> *H6: The quality of lead information in the IT-systems is positively related to the amount of information sharing between marketing and sales.*

4.3.2. Lead Prioritization & Planning

In addition to improving the quality and availability of lead information, which should create a better basis for joint activities and promote the sharing of this information, the systems also provide functions that can directly help with the prioritization and planning of lead activities. The degree to which IT-systems enable marketing and sales to assess the probability of success of marketing-generated leads and to focus and tailor their efforts on these leads is defined as support in lead prioritization and planning.

The early and appropriate identification of leads, especially of leads with higher purchase intent, has a strong impact on the conversion probability to sales (Ahearne et al., 2007; Ohiomah et al., 2019; Román & Rodríguez, 2015). Lead segmentation, scoring, and nurturing functions can help to better assess prospect's attractiveness and intent and define the best time to involve the sales rep (Mero et al., 2020; Michiels, 2009; Mrohs, 2021).

Lead scoring assigns a score to all the prospect's interactions with the company and reactions to marketing activities, as different activities may indicate a higher level of customer intent. Additionally, also basic criteria like company size, industry, or the seniority of the contact can be scored if an influence on the purchase intentions of the lead can be assumed (Mrohs, 2021; Wenger, 2021). As these criteria, scores, and thresholds for passing a lead to sales highly depend on the situation of the individual company they should be developed through marketing and sales jointly (Monat, 2011). IT-systems then offer the functionalities to manifested and automated agreed on rules and thresholds, making the designed processes more reliable and trustworthy.

Furthermore, IT-systems can help to create formalized ways to distribute leads from marketing to sales based on the scoring thresholds, and from sales back to marketing when necessary. Leads that are not ready for sales yet, can be nurtured with fitting content until they pass scoring thresholds. "Best-in-class" companies use a range of nurturing campaigns specifically designed to nurture new leads, inform prospects, reactivate inactive or closed leads, or cross-sell and up-sell to existing customers (Michiels, 2009). These functions assure that only promising leads are assigned to sales in a timely and correct manner, as the assignment and follow-up time can have a critical influence on success probabilities, especially for online leads (Oldroyd et al., 2011; Smith et al., 2006).

Nevertheless, IT-systems are only as good as the processes that are automated and defined in them. Organization cannot expect them to magically produce better processes (Michiels, 2009). This indicates that the support of IT-systems in the prioritization and planning of lead activities can help to make processes more efficient and effective, but their existence is not a sufficient condition for successful lead management. Instead, these functionalities should promote marketing and sales to actively engage in jointly determining the criteria, scores and thresholds needed to let the systems live up to their full potential. In fact, the more functionalities such systems offer the higher will be the complexity of manually setting rules and triggers (Mero et al., 2020), therefore requiring more and frequent interaction between all parties working with the system. Accordingly, the following hypothesis can be concluded:

> *H7: The IT-systems support in prioritizing and planning lead management activities is positively related to marketing and sales engagement in jointly planning lead management activities.*

The IT-systems support in the prioritization and planning of lead management activities provides both marketing and sales with valuable information about prospects and leads as it identifies and scores relevant information that can be used to assess customers purchase intent (Järvinen & Taiminen, 2016; Wenger, 2021). Based on the scores and thresholds, leads are passed to sales at the right time and marketing can consequently share more relevant information together with the lead handover. The IT-systems allow marketing to not only share basic information about a new lead but also to highlight important information they gathered like the lead score or recent activities e.g., downloads or video views (Järvinen & Taiminen, 2016; Woelke, 2021).

Properly set up systems should constantly improve the quality and informativeness of lead information in the systems (Wenger, 2021). As the systems gather as much information about customers and prospects as possible, marketing can equip sales with more accurate and detailed information about the customer's needs and purchase intent (Järvinen & Taiminen, 2016; Woelke, 2021). This allows the sales reps to better assess the quality of different leads and adjust their approaches, which should strengthen their trust in marketing's shared information and their ability to provide them only qualified leads (Ahearne et al., 2007; Moutot & Bascoul, 2008). Accordingly, as marketing is able to share more and higher quality information about the leads that are passed on to sales, the sharing of information between marketing and sales should have a stronger positive influence on the perceived quality of marketing-generated leads. Correspondingly, the following is hypothesized:

> *H8: The effect of information sharing on perceived lead quality is expected to be higher the better the IT-systems support in prioritizing and planning lead management activities is.*

5. Method

In order to analyze the research gap described above, a questionnaire was developed that addresses sales reps who regularly receive leads from their marketing department. The questionnaire contains questions regarding the constructs considered in this paper, as well as demographic questions and control variables. The sample, measures and analytic approach will be presented in more detail below.

5.1. Sample

The questionnaire was distributed to more than 1000 sales employees via LinkedIn In-Mail messages and was shared directly on LinkedIn and in different sales-specific LinkedIn groups. 495 people viewed the questionnaire, of which 159 finished it. After cleaning the data for participants not working in sales, answer biases and cases with more than 10% missing answers, a usable data set of 151 participants was obtained. The values still missing after adjustment were calculated using the Expectation Maximization (EM) method, which calculates the most probable answer based on the information provided by all other participants (Dempster et al., 1977). The Little-test (1988) indicated that missing values were missing completely at random, fulling the requirement to conduct this method. The EM method was conducted for all variables except for the control variables marketing lead volume and age, where the missing values and one outlier were replaced by the mean value because these values cannot be adequately estimated using the EM method.

Almost all the participants worked for B2B companies (B2B, N=131; B2B & B2C, N=17), only three respondents worked for B2C companies. The data set includes companies of all sizes from small and medium-sized businesses (1-99 employees; N=30), to mid-market companies (100-999 employees; N=45) and large enterprises (>1000 employees; N=74). Around one third of the respondents worked for companies with more than 5000 employees. The companies were distributed across nine different industries, with information technology & telecommunications accounting for almost two third of the sample (N=99), followed by services (N=17) and building and construction (N=9). The high proportion of companies from the information technology & telecommunications industry can probably be attributed to

the fact that these companies are more active on LinkedIn compared to other industries (Gonzalez, 2022). Potentially, it would have been useful to further divide information technology and telecommunications into different sub-industries to get a better picture of the distribution.

21.2% of the participants were female and 78.8% were male. The sample shows an age range from 21-66 years with a mean age of 38.64 (Median = 38; SD = 9.78). On average, respondents had 13.08 years of sales experience (Median = 11; SD = 9.23), had been with their current company for 4 years (Median = 2; SD = 6.12) and received 24.22 Leads per month (Median = 10; SD = 46.45).

5.2. Measures

Most of the scales used in the questionnaire were taken from previous research and adapted and extended to fit the context of this study. The questionnaire was divided into four sections, which thematically dealt with different aspects of the study.

Collaboration measures: Joint planning (JP; α=.924) was measured using a five-item scale which was adopted from the joint planning and teamwork scales of Homburg, Jensen, and Krohmer (2008) and the interfunctional coordination scale from Le Meunier-FitzHugh and Piercy (2011). The amount of information sharing (IS; α=.823) between marketing and sales was measured using a scale received from the information provision scale of Homburg, Jensen, and Krohmer (2008) which was extended by another item (IS_4 = "Marketing and sales share information about successful and unsuccessful leads fast."), based on the marketing-sales interface configurations by Biemans et al. (2010).

IT-system support measures: The quality of the lead information in the IT-systems (LI; α=.953) was measured using a six-item scale that was adapted from the perceived informativeness scale used by Choe et al. (2009) and Buaprommee and Polyorat (2016). The scale was adjusted to fit the context and extended by an item from the knowledge scale from Ahearne et al. (2007) and self-developed items. The scale for the system support in prioritization and planning (LPP; α=.94) was inspired by the customer prioritization scale used by Terho et al. (2015) and Panagopoulos and Avlonitis (2010) and the ability to assess customer profitability scale from Homburg, Droll, and Totzek (2008).

Lead management measures: For the measurement of the perceived lead quality (PLQ; α=.932) the existing scale from Sabnis et al. (2013) was used. The follow-up of marketing-generated leads (MLFU; α=.891) was measured by a four-item scale developed from two scales by Ahearne et al. (2007) and Schillewaert et al. (2005) that were modified to fit the context.

All these constructs were measured by a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree).

Control measures: It is probable that other variables, besides those hypothesized about, have influence on the model. Therefore, additional variables were included to control for their effects. These include how frequently the sales rep uses the IT-systems (ITU: IT Usage) and if their managers are tracking their follow-up activities (MT: Managerial Tracking of marketing-generated Leads). Both were measured using one-item on a seven-point-scale. In addition, the number of marketing-generated leads received by sales reps per month (MLV: Marketing Lead Volume), their age, gender, sales experience (SalYrs), company seniority (ComYrs), and company size (ComSze) were recorded.

All utilized scales and their item reliabilities are presented in the appendix in more detail (see Table 7 in the Appendix).

Reliability and validity of the scales: Internal consistency reliability was assessed using Cronbach's Alpha and Composite Reliability. As reported above, all constructs are reliable with a Cronbach's Alpha greater than .80. With Composite Reliability ranging from .841 to .951, both measures are well above the .70 benchmark (Bagozzi & Yi, 1988; Cronbach, 1951). The results are summarized in table 3.

The reliability and validity of the scales were further assessed by performing both exploratory (EFA) and confirmatory (CFA) factor analysis. The EFA extracted six factors with an Eigenvalue greater than 0.95 that explained 74.95% of the variance among the items in the study, confirming the six-dimensional structure theoretically defined (detailed results are reported in the appendix). Since the Eigenvalue of the sixth factor was 0.957, the EFA was performed forcing it to extract six factors which should be acceptable as the Kaiser's criterion has already been labeled as too strict in some cases (Field, 2009). Nonetheless, in this EFA the Item IS_1 failed to load on the right factor and therefore was removed from the analysis, which was also supported by an increase in the Cronbach's Alpha for Information Sharing (α=.823 → α = .835).

The CFA also confirmed that the six-factor model yields a great model fit for the data as all values were within their common acceptance levels (Hu & Bentler, 1999), the results can be seen in table 1.

The AVE values of all constructs were above the .50 threshold, indicating that each construct explains more than 50% of their indicator variance. Discriminant validity was assessed using the Fornell-Larcker (1981) Criterion. None of the squared correlations of the construct pairs did exceed their AVE values, therefore fulfilling the criterion (see appendix for detailed results).

5.3. Analytical Approach

The software Statistical Package for Social Science (SPSS) and the Analysis of Moment Structures (AMOS), version 26.0, were used to analyze the data. First, the reliability and validity of the scales and measurement model was tested as described above. Next, a first model (Model 1) was fit to confirm the relationship between Perceived Lead Quality (PLQ) and Follow-up of marketing-generated Leads (MLFU), and to test for indirect effects from Joint Planning (JP) and Information Sharing (IS).

After the relationship between PLQ and MLFU has been confirmed, the main effects model (Model 2) was developed

Table 1: Fit Indices for the Measurement Model

Fit Indices	Estimate	Threshold	Interpretation
CMIN	469.901	-	-
DF	306	-	-
CMIN/df	1.536	Between 1 and 3	Excellent
CFI	.956	$\geq .95$	Excellent
TLI	.949	$\geq .95$	Almost Excellent
SRMR	.0488	$\leq .05$	Excellent
RMSEA	.060	$\leq .05$	Acceptable

to test for the influence of IT-systems and to serve as a baseline model for the interaction tests. Notably, the relationship between PLQ and MLFU in this model was dropped as it has already been confirmed, and this study focuses on the predictors of PLQ. The interaction effects were measured in a third model (Model 3) that included the two interaction terms as antecedents for PLQ. To reduce potential multicollinearity effects, all variables were mean centered before the interaction terms were calculated (Dawson, 2014; Hofmann & Gavin, 1998).

6. Results

The correlations, means, and standard deviations of the variables employed in the models are displayed in table 3. As expected, there are high correlations between JP and IS, as well as between LI and LPP, as they are measuring different aspects of collaboration (JP & IS) and IT-systems (LI & LPP). The high correlations between the different variables included in the model already indicate relationships between these variables. The estimation results and model fits of all three structural models are reported in table 2.

Model 1 was designed to confirm the relationship between PLQ and MLFU, which was first described by Sabnis et al. (2013), and to test for indirect effects from JP and IS. The fit measures of this structural model, reported in table 2, indicate that the model fits the data well (Hu & Bentler, 1999). The effect of PLQ on MLFU is positive and highly significant ($\beta = 0.339$, $p < 0.01$). This offers support for H1 and further supports Sabnis et al. (2013) findings that a salespersons perceived lead quality has a significant influence on their follow-up efforts. Furthermore, a positive and significant effect from JP on PLQ was found ($\beta = 0.333$, $p < 0.01$) as well as a positive and slightly significant effect from IS on PLQ ($\beta = 0.221$, $p < 0.10$), providing support for H2 and H4. As hypothesized in H3 the results also showcase a strong positive and significant effect from IS on JP ($\beta = 0.675$, $p < 0.01$).

In addition, a mediation analysis was conducted to explore the indirect effects of IS and JP. The results are summarized in table 4. The analysis confirms a positive and significant indirect influence from both JP and IS on MLFU. PLQ fully mediates the effect of JP on MLFU as well as the effect from IS on MLFU. Additionally, it was also found that JP partially mediates the influence from IS on PLQ.

The main effects model (Model 2), now including the effects of the IT variables, was tested next. All effects already tested in Model 1 stayed positive and significant except for the effect from IS on PLQ which is now almost zero and insignificant ($\beta = 0.013$, $p > 0.10$). In support of H6, a positive and significant effect from LI on IS was found ($\beta = 0.516$, $p < 0.01$). The effect from LPP on JP also turned out to be positive and significant ($\beta = 0.236$, $p < 0.01$), providing support for H7. Additionally, a strong positive and significant effect from LI directly on PLQ ($\beta = 0.416$, $p < 0.01$) was found, which was not hypothesized.

The mediation analysis, shown in table 5, revealed that JP fully mediates the relationship between LPP and PLQ, while the indirect effect from LI on PLQ through IS turned out to be insignificant. Furthermore, it was shown that the effect from IS on PLQ is now fully mediated by JP.

Model 3 was tested last; in this model the study assessed the moderating role of LI on the relationship between JP and PLQ as well as the moderating role of LPP on the relationship between IS and PLQ. The results revealed a positive and significant moderating impact of LI on the relationship between JP and PLQ ($\beta = 0.120$, $p < 0.10$). The moderating effect of LPP on the relationship between IS and PLQ was not present and found to be insignificant ($\beta = 0.050$, $p > 0.10$). The data therefore only offers support for H5, that a higher quality of lead information in the IT-systems strengthens the effects of JP on PLQ, and H8 needs to be rejected. The interaction effect from LI on the relationship between JP and PLQ is plotted in figure 4.

A robustness check of the model was performed excluding the control variables and indicated that the results are robust (see appendix for detailed results). All effects remained significant. The only notable deviation from the Models with control variables was that the interaction effect from LI on the relationship of JP and PLQ was slightly stronger and more significant than before ($\beta = 0.143$, $p < 0.05$).

7. Discussion

This study seeks to contribute to the literature stream of lead management by presenting an empirical model that explores the influence of marketing and sales collaboration on a salesperson perceived lead quality, which can be seen as a key predictor of their lead follow-up efforts. In addition, the study also takes into account the moderating influences that

Table 2: Standardized Parameter Estimates and Model Fits

Relationship		Model 1	Model 2	Model 3
H1:	PLQ → MLFU	0.339***	-	-
	JP → MLFU	0.111	-	-
	IS → MLFU	0.198	-	-
H2:	JP→ PLQ	0.333***	0.256***	0.254***
H3:	IS → JP	0.675***	0.564***	0.564***
H4:	IS → PLQ	0.221***	0.013	0.030
H5:	LI × JP → PLQ	-	-	0.120*
H6:	LI → IS	-	0.516***	0.516***
H7:	LPP → JP	-	0.236***	0.235***
H8:	LPP × IS → PLQ	-	-	0.050
	LI → PLQ	-	0.436***	0.417***
	LPP → PLQ	-	0.154	0.168*
Control variables:				
	MLV → PLQ	-.084	-.051	-.048
	ITU → PLQ	0.188**	0.029	0.028
	MT → PLQ	0.098	-.013	-.013
	Gender → PLQ	0.037	0.070	0.058
	ComSze → PLQ	0.103	0.009	-.009
	ComYrs → PLQ	0.000	0.043	0.040
	SalYrs → PLQ	-.061	-.026	-.031
	Age → PLQ	0.053	0.074	0.076
Model Fits:				
CMIN (df)		299.856 (201)	517.070 (379)	611.061 (419)
CMIN/df		1.492	1.478	1.458
RMSEA		.057	.057	.056
SRMR		.054	.056	.055
CFI		.953	.949	.947
TLI		.936	.938	.933

= Significant at p<.10; **= Significant at p<.05; *= Significant at p<0.01*

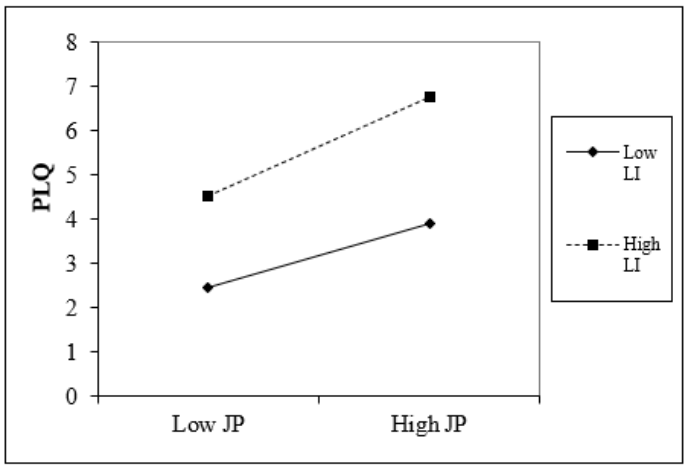


Figure 4: The Moderating Role of LI on the JP-PLQ Relationship

IT-systems can have on these relationships as well as how IT-systems could increase marketing and sales collaboration. All this should help in revealing how marketing and sales managers can best set up their lead management programs for success.

The study proposed eight different hypotheses and found support for six of them, an overview of the results can be found in table 6.

Table 3: Correlations, Means, and Standard Deviation of the Used Variables

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Joint Planning														
2. Information Sharing	.595**													
3. Quality of Lead Information	.395**	.464**												
4. Support in Lead Prioritization & Planning	.427**	.334**	.684**											
5. Perceived Lead Quality	.494**	.412**	.655**	.585**										
6. Lead Follow-up	.366**	.362**	.461**	.432**	.492**									
Control Variables														
7. IT Usage	.233**	.192*	.383**	.376**	.314**	.271**								
8. Marketing Lead Volume	.215**	.152	.034	.072	.026	.033	.023							
9. Managerial Tracking	.225**	.112	.317**	.309**	.238**	.260**	.205*	.141						
10. Gender	.017	.023	-.018	.069	.061	.145	.036	-.004	.005					
11. Age	.103	-.006	-.165*	-.167*	-.051	.065	-.160	-.021	-.088	-.185*				
12. Company Size	-.185*	-.145	.015	.019	-.029	-.077	-.229**	-.110	-.071	.084	-.184*			
13. Company Seniority	-.199*	-.127	-.240**	-.274**	-.164*	-.226**	-.114	-.091	-.188*	-.048	.271**	.009		
14. Sales Experience	.043	.015	-.169*	-.181*	-.078	-.051	-.145	-.054	-.070	-.223**	.806**	-.117	.383**	
Mean	4.52	4.64	4.35	4.21	3.48	5.10	6.23	24.99	4.42	1.21	38.64	4.08	4.47	13.08
Standard Deviation	1.58	1.51	1.67	1.69	1.69	1.56	1,16	46.45	1.95	.41	9.78	1.83	6.12	9.23
Cronbach's Alpha (CrA)	.924	.823	.953	.941	.932	.890								
Composite Reliability (CR)	.921	.841	.951	.941	.937	.894								
Average Variance Extracted (AVE)	.700	.640	.764	.760	.788	.680								

**= Significant at p<.05 (two-tailed); **= Significant at p<.01 (two-tailed)*

Table 4: Mediation Analysis - Model 1

Predictor (X)	Mediator (M)	Outcome (Y)	X→ M	M → Y	X → Y	Indirect Effect
JP	PLQ	MLFU	.333***	.339***	.111	.113**
IS			.221*		.198	.075*
IS	JP	PLQ	.675***	.333***	.221*	.225**

Note: Standardized Estimates are reported
= Significant at p<.10; **= Significant at p<.05; *= Significant at p<0.01*

Table 5: Mediation Analysis - Model 2

Predictor (X)	Mediator (M)	Outcome (Y)	X→ M	M → Y	X → Y	Indirect Effect
LI	IS	PLQ	.516***	.013	.436***	.007
LPP	JP	PLQ	.236***	.256***	.154	.060**
IS	JP	PLQ	.564***	256***	.013	.144**

Note: Standardized Estimates are reported
= Significant at p<.10; **= Significant at p<.05; *= Significant at p<0.01*

Table 6: Summary of Results

Hypothesis	Independent Variable	Dependent Variable	Moderator	Results
H1	PLQ	MLFU	-	✓
H2	JP	PLQ	-	✓
H3	IS	JP	-	✓
H4	IS	PLQ	-	∼
H5	JP	PLQ	LI	✓
H6	LI	IS	-	✓
H7	LPP	JP	-	✓
H8	IS	PLQ	LPP	✗

✓= supported; ✗= not supported; ∼ = partially supported

7.1. Research Issues

The predictors of perceived lead quality have remained widely unexplored as research has mainly focused on options to determine the objective quality of leads (e.g., D'Haen and van den Poel, 2013). This paper addressed this research gap by examining the influence of one of the most important factors in lead management, the collaboration between marketing and sales, while also considering potential effects the support from IT-systems may have. With a sample of 151 salespersons that regularly work with marketing-generated leads, this study gained interesting insights on the interplay between IT-systems and collaboration in influencing the perceived quality of marketing-generated leads.

First, the results of this paper were able to further strengthen the findings from Sabnis et al. (2013), that the perceived quality of marketing-generated leads has a significant influence on a sales rep's follow-up effort. The better a salesperson's perception of marketing-generated leads gets, the more willing they are to follow-up on these leads. This fact already stresses the importance of a good prequalification process for marketing-generated leads that needs to match sales expectations.

The beneficial effects of marketing and sales collaboration have already been discussed in interface literature but have not been transferred on lead management research yet. The results of this study showcase that collaboration has an important impact on the lead management process, as it was confirmed as an important predictor of the perceive quality of marketing-generated leads. Therefore, also influencing the follow-up efforts of sales reps through indirect relationships.

Joint Planning showcased the strongest effect on the perceived quality of marketing-generated leads among the collaboration variables. The more and the closer marketing and sales plan lead management activities and processes together, the higher is the perceived lead quality of the sales reps. This can be explained by a higher level of commitment and trust in the jointly developed processes, as it is ensured that the perspectives and requirements of both parties are taken into account. The influence from information sharing on perceived lead quality was weaker and only remained significant as long as the IT variables were not considered. The insignificance of the effect within the full model may be explained by the fact that, if the IT-systems provide sales and marketing with high-quality and well-organized information about marketing-generated leads, the beneficial effect of addition-

ally sharing this information becomes very small. The not hypothesized strong direct effect from quality of lead information on the perceived lead quality supports this interpretation.

IT-systems were also confirmed to play an important role in lead management processes. It was found that different aspects of IT-systems can significantly influence the collaboration between marketing and sales. The quality of the lead information in the IT-systems was found to positively influence the amount of information sharing between marketing and sales. This can be explained by the fact, that high-quality and well-organized IT-systems decrease the thresholds to share information by making information easier to access and share, while also highlighting the value and necessity of sharing relevant information. Furthermore, IT-systems support in the prioritization and planning of lead management activities was found to positively influence the amount of joint planning between marketing and sales. As IT-systems cannot magically create or improve lead management processes, it is necessary that marketing and sales consult about the setup of such systems and therefore also about the lead management process in general. This is an important contribution, as it demonstrates the potential of IT-systems in aligning marketing and sales.

IT-systems are not only able to improve the collaboration between marketing and sales but were also found to strengthen the positive effect of collaboration on the perceived quality of marketing-generated leads. A high quality of lead information in the IT-systems creates a better foundation for marketing and sales planning activities. Decisions can be made faster and are based on better and more relevant data, creating more confidence in agreed processes. This finding further strengthens companies need for high quality IT-systems and stresses the importance of good data management and the integration of different systems in a single source of truth. No evidence was found that IT-systems support for prioritization and planning had a strengthening effect on the relationship between information sharing and the perceived quality of leads generated by marketing. This might be attributed to the same reason as the reason for the missing relationship between information sharing and perceived quality of marketing-generated leads in the model with IT variables.

It can be concluded that both collaboration between marketing and sales and IT-systems play an important role in lead management and are essential for building a good perception of marketing-generated leads. Further can be noted that IT-systems seem to enable lead management in its best possible form as they not only strongly increase the collaboration between marketing and sales, but also enhance the effects that this collaboration has on the perceived quality of marketing-generated leads. All in all, there is a complex interplay between collaboration and IT-systems, which implies that only successful initiatives in both areas promise a maximum of success.

7.2. Managerial Implications

Each year, a large portion of marketing budgets is spent on the creation of new leads, increasingly through online marketing channels (Gartner, 2021). If marketing generates leads and sales does not follow up on them, a lot of valuable resources are wasted. The value of this study for exploring ways to improve sales reps' response rates to marketing-generated leads is therefore obvious. Valuable implications on how to design lead management processes to reduce the waste of resources can be derived from this study.

First, evidence has been found that the joint planning of lead management activities has an important impact on salespersons' perception of lead quality. This illustrates that lead management must be seen as neither a marketing nor a sales task, but as a joint process in which both parties must be equally involved. When marketing and sales coordinate and plan their processes together, both parties are more committed to the process and know how to address and resolve issues in the process. This ensures that the different expectations and requirements that marketing and sales may have on a lead, or the lead management process, are factored in and that everyone knows what to expect.

Information sharing was also found to have a positive impact on the perceived quality of leads generated by marketing. However, as this influence fades once IT-systems are included in the model, it is recommended that sales and marketing managers should focus more on implementing high-quality IT-systems instead. Systems with high-quality and well-organized data were found to increase the amount of information sharing between marketing and sales. In addition, the functions of IT-systems that help prioritize and plan lead activities encourage marketing and sales to plan processes and activities together to a greater extent.

Furthermore, the study stresses the importance of high-quality and well-organized data within the IT-systems. A high quality of lead information in the systems was shown to significantly enhance the effects that joint planning has on the perceived lead quality by creating a better foundation for joint decisions. On top of that, it also had one of the strongest direct effects on the perceived lead quality. These results show the great benefits of good data management. It should prompt marketing and sales managers to ensure that the most important data about leads is kept up to date and presented in a way that allows everyone to access key information quickly and easily. Especially the integration of marketing and sales systems should be of great benefit in this, as recommended by others before (see e.g., Wiersema, 2013).

7.3. Limitations, Conclusions and Directions for Future Research

There are some limitations to this study that can restrict its generalizability and interpretation. These are discussed in the following section, including possible directions for future research.

One limitation can be seen in the fact that the study only measured data on the salesperson level. Therefore, not only

the perception of lead quality, but also the collaboration and IT variables were measured solely on the subjective level of the salesperson. This can be justified with the studies focus on the subjective quality of leads, which of course should be most influenced by a salesperson's subjective perception of collaboration and IT-systems. Nevertheless, measuring the data also from the marketing perspective may generate further insights in the processes and could be considered in future research.

Another limitation resulting from the one level measurement is that the follow-up efforts were self-reported by the salespersons. This means that the self-reported data may be biased, and results would have been more robust if the follow-up efforts had been measured on the sales manager level or through actual CRM-Data about lead management activities (see e.g., Gramzow et al., 2003).

As reported in the description of the sample, around two third of the respondents worked in the information technology & telecommunications industry. Due to the high percentage of respondents from the same industry, it might be possible that the sample is biased, and certain effects only arose due to specific policies especially applied in this industry. Therefore, the generalizability of the results to other industries may be restricted and needs to be interpreted with caution. Future research that replicates the results of this study, while focusing on other industries, could help to further strengthen and validate the findings.

Collaboration and IT-systems were both assessed via two different aspects. It is obvious that there are more facets of collaboration and IT-systems, and therefore likely more ways how collaboration and IT-systems can influence each other and the perceived quality of marketing-generated leads. Examples of other influencing factors could include knowledge about the other parties' day-to-day activities, the quality of the exchange and interaction between both parties, or the user-friendliness of IT-systems. In addition, it may be interesting to further investigate which specific types of IT-systems, e.g., sales force automation or marketing automation, contribute most to successful lead management. All these points might be interesting to consider in future research.

Despite the above limitations, this study makes an important contribution to the literature on lead management and paves the way for further research in this area. It is demonstrated that both collaboration between marketing and sales and IT-systems make a significant contribution to successful lead management, and that the successful combination of both open up opportunities to overcome the sales lead black hole.

References

Adamson, B., Dixon, M., & Toman, N. (2012). The end of solution sales. *Harvard Business Review*, *90*(7–8), 60–68.

Ahearne, M., Hughes, D. E., & Schillewaert, N. (2007). Why sales reps should welcome information technology: Measuring the impact of CRM-based IT on sales effectiveness. *International Journal of Research in Marketing*, *24*(4), 336–349.

Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94.

Banerjee, S., & Bhardwaj, P. (2019). Aligning marketing and sales in multichannel marketing: Compensation design for online lead generation and offline sales conversion. *Journal of Business Research*, *105*, 293–305.

Beverland, M., Steel, M., & Dapiran, G. P. (2006). Cultural frames that drive sales and marketing apart: an exploratory study. *Journal of Business & Industrial Marketing*, *21*(6), 386–394.

Bhattacharyya, S. (2014). Improving inventory demand forecasting by using the sales pipeline: A case study. *The Journal of Business Forecasting*, *33*(1), 7.

Biemans, W. G., Makovec Brenčič, M., & Malshe, A. (2010). Marketing–sales interface configurations in B2B firms. *Industrial Marketing Management*, *39*(2), 183–194.

Bradford, W. R., Johnston, W. J., & Bellenger, D. (2016). The Impact of Sales Effort on Lead Conversion Cycle Time in a Business-to-Business Opportunity Pipeline. *SSRN Journal*.

Buaprommee, N., & Polyorat, K. (2016). Intention to purchase traceable meat: The impacts of perceived information asymmetry, informativeness, usefulness, and norm. *Asian Journal of Business and Accounting*, *9*(1), 141–168.

Choe, Y. C., Park, J., Chung, M., & Moon, J. (2009). Effect of the food traceability system for building trust: Price premium and buying behavior. *Information Systems Frontiers*, *11*(2), 167–179.

Colletti, J. A., & Chonko, L. B. (1997). Change management initiatives: moving sales organizations from obsolescence to high performance. *Journal of Personal Selling & Sales Management*, *17*(2), 1–30.

Collins, R. H. (1985). Microcomputer systems to handle sales leads: A key to increased salesforce productivity. *Journal of Personal Selling & Sales Management*, *5*(1), 77–83.

Cooper, M. J., & Budd, C. S. (2007). Tying the pieces together: A normative framework for integrating sales and project operations. *Industrial Marketing Management*, *36*(2), 173–182.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Dawes, P. L., & Massey, G. R. (2005). Antecedents of conflict in marketing's cross-functional relationship with sales. *European Journal of Marketing*.

Dawson, J. F. (2014). Moderation in Management Research: What, Why, When, and How. *Journal of Business and Psychology*, *29*(1), 1–19.

Deci, E. L., & Ryan, R. M. (2000). The "What" and "Why" of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, *11*(4), 227–268.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

D'Haen, J., van den Poel, D., Thorleuchter, D., & Benoit, D. F. (2016). Integrating expert knowledge and multilingual web crawling data in a lead qualification system. *Decision Support Systems*, *82*, 69–78.

D'Haen, J., & van den Poel, D. (2013). Model-supported business-to-business prospect prediction based on an iterative customer acquisition framework. *Industrial Marketing Management*, *42*(4), 544–551.

Engle, R. L., & Barnes, M. L. (2000). Sales force automation usage, effectiveness, and cost-benefit in Germany, England and the United States. *Journal of Business & Industrial Marketing*.

Field, A. (2009). *Discovering Statistics Using SPSS. (and sex and drugs and rock'n'roll)* (3rd ed.). Sage.

Fisher, R. J., Maltz, E., & Jaworski, B. J. (1997). Enhancing communication between marketing and engineering: The moderating role of relative functional identification. *Journal of Marketing*, *61*(3), 54–70.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39–50.

Fröhlich, E. (2021). Diese B2B Trends kommen 2022 im Online-Marketing auf dich zu. Retrieved May 8, 2022, from https://www.eminded.de/blog/b2b-trends-online-marketing/

Gartner. (2021). The State of Marketing Budgets 2021.

Gonzalez, S. (2022). LinkedIn Industries List & Rankings 2022. Retrieved May 8, 2022, from https://sage.exchange/linkedin-industries-list-rankings-2022/

Gramzow, R. H., Elliot, A. J., Asher, E., & McGregor, H. A. (2003). Self-evaluation bias and academic performance: Some ways and some reasons why. *Journal of Research in Personality*, *37*(2), 41–61.

Guenzi, P., & Troilo, G. (2006). Developing marketing capabilities for customer value creation through Marketing–Sales integration. *Industrial Marketing Management*, *35*(8), 974–988.

Guenzi, P., & Troilo, G. (2007). The joint contribution of marketing and sales to the creation of superior customer value. *Journal of Business Research*, *60*(2), 98–107.

Hannig, U., Heinzelbecker, K., & Foell, T. (2019). Digital Safety First. Marketing Automation in DACH 2019.

Hasselwander, A. (2006). B2B Pipeline Management. Retrieved May 8, 2022, from https://www.b2bmarketingconfidential.com/b2b-pipeline-management/

Hise, R. T., & Reid, E. L. (1994). Improving the performance of the industrial sales force in the 1990s. *Industrial Marketing Management*, *23*(4), 273–279.

Hofmann, D. A., & Gavin, M. B. (1998). Centering Decisions in Hierarchical Linear Models: Implications for Research in Organizations. *Journal of Management*, *24*(5), 623–641.

Homburg, C., Droll, M., & Totzek, D. (2008). Customer prioritization: does it pay off, and how should it be implemented? *Journal of Marketing*, *72*(5), 110–130.

Homburg, C., & Jensen, O. (2007). The Thought Worlds of Marketing and Sales: Which Differences Make a Difference? *Journal of Marketing*, *71*(3), 124–142.

Homburg, C., Jensen, O., & Krohmer, H. (2008). Configurations of Marketing and Sales: A Taxonomy. *Journal of Marketing*, *72*(2), 133–154.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1–55.

HubSpot. (2022a). Signavio Generates 188% More Leads in a Year. Retrieved May 8, 2022, from https://www.hubspot.com/case-studies/signavio

HubSpot. (2022b). Warum HubSpot? Die CRM-Plattform von HubSpot. Retrieved May 8, 2022, from https://www.hubspot.com/de/why-hubspot

Hulland, J., Nenkov, G. Y., & Barclay, D. W. (2012). Perceived marketing–sales relationship effectiveness: a matter of justice. *Journal of the Academy of Marketing Science*, *40*(3), 450–467.

Hunter, G. K., & Perreault Jr., W. D. (2006). Sales Technology Orientation, Information Effectiveness, and Sales Performance. *Journal of Personal Selling & Sales Management*, *26*(2), 95–113.

Järvinen, J., & Taiminen, H. (2016). Harnessing marketing automation for B2B content marketing. *Industrial Marketing Management*, *54*, 164–175.

Jelinek, R., Ahearne, M., Mathieu, J., & Schillewaert, N. (2006). A longitudinal examination of individual, organizational, and contextual factors on sales technology adoption and job performance. *Journal of Marketing Theory and Practice*, *14*(1), 7–23.

Jolson, M. A. (1988). Qualifying sales leads: The tight and loose approaches. *Industrial Marketing Management*, *17*(3), 189–196.

Jolson, M. A., & Wotruba, T. R. (1992). Selling and Sales Management in Action: Prospecting: A New Look at this Old Challenge. *Journal of Personal Selling & Sales Management*, *12*(4), 59–66.

Kahn, K. B., & Mentzer, J. T. (1998). Marketing's integration with other departments. *Journal of Business Research*, *42*(1), 53–62.

Kirca, A. H., Jayachandran, S., & Bearden, W. O. (2005). Market orientation: A meta-analytic review and assessment of its antecedents and impact on performance. *Journal of Marketing*, *69*(2), 24–41.

Kotler, P., Rackham, N., & Krishnaswamy, S. (2006). Ending the war between sales and marketing. *Harvard business review*, *84*(July/August), 67–78.

Kuruzovich, J. (2013). Sales technologies, sales force management, and online infomediaries. *Journal of Personal Selling & Sales Management*, *33*(2), 211–224.

Le Meunier-FitzHugh, K., & Piercy, N. F. (2011). Exploring the Relationship between Market Orientation and Sales and Marketing Collaboration. *Journal of Personal Selling & Sales Management*, *31*(3), 287–296.

Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Long, M. M., Tellefsen, T., & Lichtenthal, J. D. (2007). Internet integration into the industrial selling process: A step-by-step approach. *Industrial Marketing Management*, *36*(5), 676–689.

Malshe, A., Johnson, J. S., & Viio, P. (2017). Understanding the sales-marketing interface dysfunction experience in business-to-business firms: A matter of perspective. *Industrial Marketing Management*, *63*, 145–157.

Malshe, A., & Sohi, R. S. (2009a). Sales Buy-In of Marketing Strategies: Exploration of its Nuances, Antecedents, and Contextual Conditions. *Journal of Personal Selling & Sales Management*, *29*(3), 207–225.

Malshe, A., & Sohi, R. S. (2009b). What makes strategy making across the sales-marketing interface more successful? *Journal of the Academy of Marketing Science*, *37*(4), 400–421.

Matthyssens, P., & Johnston, W. J. (2006). Marketing and sales: optimization of a neglected relationship. *Journal of Business & Industrial Marketing*, *21*(6), 338–345.

Mero, J., Tarkiainen, A., & Tobon, J. (2020). Effectual and causal reasoning in the adoption of marketing automation. *Industrial Marketing Management*, *86*, 212–222.

Michiels, I. (2009). Lead lifecycle management: Building a pipeline that never leaks.

Monat, J. P. (2011). Industrial sales lead conversion modeling. *Marketing Intelligence & Planning*, *29*(2), 178–194.

Moran, M. (2022). 30+ Lead Generation Statistics, Facts & Trends For 2022. Retrieved May 8, 2022, from https://startupbonsai.com/lead-generation-statistics/

Moutot, J.-M., & Bascoul, G. (2008). Effects of sales force automation use on sales force activities and customer relationship management processes. *Journal of Personal Selling & Sales Management*, *28*(2), 167–184.

Mrohs, A. (2021). Marketing Automation: Defining the Organizational Framework. In U. G. Seebacher (Ed.), *B2B Marketing* (pp. 211–227). Springer International Publishing.

Ohiomah, A., Andreev, P., Benyoucef, M., & Hood, D. (2019). The role of lead management systems in inside sales performance. *Journal of Business Research*, *102*, 163–177.

Ohiomah, A., Benyoucef, M., & Andreev, P. (2016). Driving inside sales performance with lead management systems: A conceptual model. *Journal of Information Systems Applied Research*, *9*(1), 4.

Oldroyd, J. B., McElheran, K., & Elkington, D. (2011). The short life of online sales leads. *Harvard Business Review*, *89*(3), 28.

Oliva, R. A. (2006). The three key linkages: Improving the connections between marketing and sales. *Journal of Business & Industrial Marketing*, *21*(6), 395–398.

Panagopoulos, N. G., & Avlonitis, G. J. (2010). Performance implications of sales strategy: The moderating effects of leadership and environment. *International Journal of Research in Marketing*, *27*(1), 46–57.

Park, J. E., Kim, J., Dubinsky, A. J., & Lee, H. (2010). How does sales force automation influence relationship quality and performance? The mediating roles of learning and selling behaviors. *Industrial Marketing Management*, *39*(7), 1128–1138.

Patterson, L. (2007). Marketing and sales alignment for improved effectiveness. *Journal of Digital Asset Management*, *3*(4), 185–189.

Peterson, R. M., Rodriguez, M., & Krishnan, V. (2011). CRM and sales pipeline management: Empirical results for managing opportunities. *Marketing Management Journal*, *21*(1), 60–70.

Plouffe, C. R., Williams, B. C., & Leigh, T. W. (2004). Who's on first? Stakeholder differences in customer relationship management and the elusive notion of "shared understanding". *Journal of Personal Selling & Sales Management*, *24*(4), 323–338.

Pullins, E. B. (2001). An exploratory investigation of the relationship of sales force compensation and intrinsic motivation. *Industrial Marketing Management*, *30*(5), 403–413.

Pullins, E. B., Timonen, H., Kaski, T., & Holopainen, M. (2017). An investigation of the theory practice gap in professional sales. *Journal of Marketing Theory and Practice*, *25*(1), 17–38.

Rapp, A., Agnihotri, R., & Forbes, L. P. (2008). The sales force technology–performance chain: The role of adaptive selling and effort. *Journal of Personal Selling & Sales Management*, *28*(4), 335–350.

Román, S., & Rodríguez, R. (2015). The influence of sales force technology use on outcome performance. *Journal of Business & Industrial Marketing*, *30*(6), 771–783.

Rouziès, D., Anderson, E., Kohli, A. K., Michaels, R. E., Weitz, B. A., & Zoltners, A. A. (2005). Sales and marketing integration: A proposed framework. *Journal of Personal Selling & Sales Management*, *25*(2), 113–122.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54–67.

Sabnis, G., Chatterjee, S. C., Grewal, R., & Lilien, G. L. (2013). The Sales Lead Black Hole: On Sales Reps' Follow-Up of Marketing Leads. *Journal of Marketing*, *77*(1), 52–67.

Schillewaert, N., Ahearne, M. J., Frambach, R. T., & Moenaert, R. K. (2005). The adoption of information technology in the sales force. *Industrial Marketing Management*, *34*(4), 323–336.

Schmitz, C., Altenscheidt, J., Ernst, S. C., & Huckemann, M. (2020). Digitalisierung und Automatisierung, Herausforderungen und Gestaltungsansätze für eine kundenzentrierte Marktbearbeitung. *Marketing Review St. Gallen*, 18–26.

Schrage, M. (1990). *Shared Minds: The New Technologies of Collaboration*. Random House.

Siguaw, J. A., Brown, G., & Widing, R. E. (1994). The influence of the market orientation of the firm on sales force behavior and attitudes. *Journal of Marketing Research*, *31*(1), 106–116.

Smith, T. M., Gopalakrishna, S., & Chatterjee, R. (2006). A Three-Stage Model of Integrated Marketing Communications at the Marketing-Sales Interface. *Journal of Marketing Research*, *43*(4), 564–579.

Souder, W. E. (1987). *Managing New Product Innovations*. Lexington Books.

Speier, C., & Venkatesh, V. (2002). The Hidden Minefields in the Adoption of Sales Force Automation Technologies. *Journal of Marketing*, *66*(3), 98–111.

Strahle, W. M., Spiro, R. L., & Acito, F. (1996). Marketing and sales: strategic alignment and functional implementation. *Journal of Personal Selling & Sales Management*, *16*(1), 1–20.

Tanner, J. F., Ahearne, M., Leigh, T. W., Mason, C. H., & Moncrief, W. C. (2005). CRM in sales-intensive organizations: A review and future directions. *Journal of Personal Selling & Sales Management*, *25*(2), 169–180.

Team Linchpin. (2022). Projected 2022 Average Cost Per Lead By Industry And Channel. Retrieved May 8, 2022, from https://linchpinseo.com/average-cost-per-lead/

Terho, H., Eggert, A., Haas, A., & Ulaga, W. (2015). How sales strategy translates into performance: The role of salesperson customer orientation and value-based selling. *Industrial Marketing Management*, *45*, 12–21.

Torkornoo, S. (2020). Sales und Marketing Herausforderungen 2020. SiversGroup.

van der Borgh, M., Xu, J., & Sikkenk, M. (2020). Identifying, analyzing, and finding solutions to the sales lead black hole: A design science approach. *Industrial Marketing Management*, *88*, 136–151.

Wenger, S. (2021). Successful Lead Management: Nothing's Gonna Stop Us Now. In U. G. Seebacher (Ed.), *B2B Marketing* (pp. 247–269). Springer International Publishing.

Wiersema, F. (2013). The B2B agenda: The current state of B2B marketing and a look ahead. *Industrial Marketing Management*, *42*(4), 470–488.

Woelke, A. (2021). Zusammenspiel von Lead Management und CRM entlang der Customer Journey. In U. Hannig (Ed.), *Marketing und Sales Automation* (pp. 215–223). Springer Fachmedien Wiesbaden.

Yan, J., Gong, M., Sun, C., Huang, J., & Chu, S. M. (2015). Sales pipeline win propensity prediction: A regression approach. *2015 IFIP/IEEE International Symposium on Integrated Network Management (IM)*, 854–857.

Zoltners, A. A., Sinha, P., & Zoltners, G. A. (2001). *The complete guide to accelerating sales force performance*. AMACOM/American Management Association.

Junior Management Science 9(3) (2024) 1700-1732

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Accelerator Impact on Peer Networking - Examining the Formation, Use, and Development of Inter-Organizational Networks Among Early-Stage Start-Ups

Hannah Franziska Gundel

Technical University of Munich

Abstract

Developing, promoting, and managing networks is a core element of entrepreneurship. Yet, limited research exists on the inherent processes and interaction dynamics underlying the social phenomenon of network formation among nascent companies over time. I conducted a qualitative inductive study with ten founding teams over three months to gain new theoretical insights into inter-organizational network formation among early-stage start-ups in an accelerator environment. The systematically derived dynamic process model proposes a framework capturing different types of peer relationships that change in response to founders' shifting personal and organizational needs over time. It highlights the accelerator's intermediary role in orchestrating network formation among founders through strategic design choices and regulatory program structure, establishing a collectivist organizational culture. Findings point to the entrepreneur's particular context in identifying relevant collaboration opportunities and navigating effective start-up networks, significantly informing the entrepreneurial career trajectory. The theoretical framework offers guidance for ecosystem builders, policy makers as well as opens possibilities for further research in social science and the entrepreneurial landscape.

Keywords: accelerator; entrepreneurial networking; inter-organizational networks; network orchestration; peer networks

1. Introduction

Entrepreneurs are commonly seen as autonomous business owners who develop innovative products or services and pursue new business opportunities, often aggressively and driven by personal interests (Engel, Kaandorp, & Elfring, 2016). While this archetype of the entrepreneur may be true for some individuals, there is much more to entrepreneurship and the individual entrepreneur. What is often overlooked in reality is the intersection between the individual entrepreneurs and the environment surrounding them (Acs et al., 2014; Adner & Kapoor, 2010). Entrepreneurs are embedded in a social network composed of a multitude of interdependencies between all actors in and around the network. Accordingly, not only social networks as a whole but, in particular, the individual sub-relationships within the network as well as within and between members of the network play an integral role in the viability and success of nascent companies (Engel et al., 2017).

1.1. Social networks in entrepreneurship

Although the general value of networks has long been recognized across the entrepreneurial landscape, much of the relevant literature continues to focus on the study of a network's diverse architecture and underlying structural properties (Engel et al., 2017). However, emerging research interest in the activities of entrepreneurs to shape their network relationships underlines the effort of scholars to move away from traditional deterministic approaches of tie formation (see Hoang and Antoncic, 2003; Porter and Woo, 2015; Stuart and Sorenson, 2007). In line with recent developments, entrepreneurs are no longer understood as passive nodes inside rigid network structures, but as self-determined agents who actively and consciously shape their individual community (Hallen & Eisenhardt, 2012; Vissa, 2012). Along with this agent-oriented view, researchers hypothesize that the individual entrepreneur tends to act strategically in forming efficient ties to gain access to resources, create beneficial partnerships, and discover novel opportunities (Hallen &

DOI: https://doi.org/10.5282/jums/v9i3pp1700-1732

Eisenhardt, 2012; Stuart & Sorenson, 2007; Vissa, 2012). Although at first glance, this seems to be a plausible theoretical approach, critical voices point to a number of related operational shortcomings: First of all, this approach reflects an outdated form of heroic behavior on the part of the entrepreneur. Second, entrepreneurial activity is associated with a high degree of outcome uncertainty (Alvarez & Barney, 2005; Burns et al., 2016; McMullen & Shepherd, 2006). In other words, start-up processes are accompanied by unpredictable events, goal ambiguity, and a highly volatile environment, making purely strategic network design increasingly challenging (Alvarez & Barney, 2007; Burns et al., 2016; Miller, 2007). Finally, early-stage start-ups and nascent founders, in particular, often lack the initial network contacts needed to pursue strategically targeted connections in the first place (Engel et al., 2017). Hence, identifying and entering into business-relevant relationships and building strategic social networks is arguably one of the biggest challenges for founding teams, particularly in the early stages of development.

1.2. Peer networks in the accelerator context

To overcome these particular difficulties faced by young entrepreneurs, numerous business support institutions have emerged over the past decade in an attempt to bridge the gap between nascent founders and the broader regional community. Beyond their ability to access a strong external network of universities, investors, and government institutions, a growing body of literature emphasizes the value of internal network structures and relationships taking place within these institutions. In this context, scholars point to the significant role of collaborative networks, whose inherent social interactions and mechanisms may have strong influence on the development of entrepreneurial ventures (Bøllingtoft & Ulhøi, 2005; Krishnan et al., 2020; Soetanto & Jack, 2013).

Facilitating social connections and building an internal community of founders is typically viewed as an essential competency relevant to the nascent phenomenon of start-up accelerators (Hallen et al., 2020). Because the accelerator is not only strongly connected to the broader regional ecosystem but also forms the center of its internal network, it seems to have a crucial role in influencing internal network structures and connectivity among founders (Soetanto & Jack, 2013). In addition, the intense cohort experience particularly evident in accelerator environments is perceived as an excellent opportunity for young founders to connect with peers and learn from the shared experiences of similar individuals (Hallen et al., 2020). Yet, relationships and networks between start-ups prove to be a very abstract and fragmented phenomenon within entrepreneurship research, particularly due to their fluid and dynamic structures associated with starting a new business (Jack et al., 2008).

1.3. Relevance of the research

In this regard, previous literature has conceptualized entrepreneurial networks from a relatively static perspective attempting to explain entrepreneurial behavior based on individual network characteristics. In other words, academic research has been primarily devoted to unlocking the structural components such as individual characteristics and goals of network building (i.e., the *what*), while relatively little is known about the procedural elements such as behaviors and processes (i.e., the *why* and *how*) underlying entrepreneurial networking (Evald et al., 2006). For example, we know little about the social mechanisms that may improve the identification and use of valuable connections in a network. Do early-stage founders find their way to relevant contacts on their own, or do they rely on external support to help them build relationships with other founders? Are memberships in entrepreneurial social networks likely to create a sense of belonging or rivalry among peers? How do relationships relate to the dynamics of the entrepreneurial process, and how do the different forms of connections evolve over time? Due to the lack of research in unlocking the dynamic components of entrepreneurial networks, scholars are increasingly pointing to the need for process-oriented research as an empirical approach to developing a more comprehensive and in-depth understanding of the entrepreneurial phenomenon (McMullen & Dimov, 2013). For this reason, the focus should shift from only examining the characteristics of particular types of relationships to the associated effects and longevity of these different forms (Aldrich & Zimmer, 1986).

Hence, to truly understand the dynamic processes of network formation and the associated social interaction mechanisms between start-ups, it is required to transfer the study of social networks to an entrepreneurial setting, locate it in the unit of analysis between emergence and development (Davidsson & Wiklund, 2001), and follow a holistic approach by examining the contextual phenomenon over time (Dimov, 2011).

1.4. Research objectives and research questions

Accordingly, this paper adopts a qualitative, inductive approach to provide information on the dynamics and change processes of relationship formation over time. The goal of this paper is, therefore, to shed light on the phenomenon of relationship formation between young founders and founding teams to develop a comprehensive understanding of the governing characteristics and dynamics in this process. The process-oriented view allows me to include the temporal context of network formation and draw theoretical conclusions about how the relationships between founders evolve from the first encounter to the end of the study period (Langley, 1999). Because entrepreneurial relationships unfold and underlie the dynamics of a continuous process, a suitable setting is necessary in order to overcome this particular limitation. Since the accelerator environment provides a social system that enables the formation of networks among early-stage entrepreneurs, it seems to be a particularly favorable research setting for observing the phenomenon under study. So far, academic literature has mainly examined accelerator networks in terms of what components they are built of (see Cohen, Fehder, et al., 2019). This paper expands on these findings to discuss how they emerge, change and sustain to understand the impact social networks may have in terms of

accelerators' effectiveness. It further illuminates the role of the accelerator in channeling collective experiences between its tenants. In this regard, I seek to answer the following research questions:

> *(1) How do early-stage start-ups form, use, and develop relationships with other founders?*
>
> *(2) How does the importance of different relationships change in light of the dynamic nature of entrepreneurial activity?*
>
> *(3) How does the involvement of accelerators impact the process of relationship building between founders?*

1.5. Data set and methodology

In line with the previously outlined research questions, I seek to understand the dynamic logic behind the development of relationships among start-ups in their early stages. Thus, since the analysis focuses on the process of relationship building and its related dimensions, I adopt a qualitative inductive approach involving semi-structured interviews with ten early-stage start-ups. In order to gain a comprehensive and transparent insight into the phenomenon under study, I choose a context that promises rich data and a high degree of explanatory power across all individual process steps associated with it (Patton, 2002; Yin, 1994). The accelerator setting provides a natural environment to observe the emergence, change and development of inter-firm[1] relationships over an extended period of time. The approach of a single case study seems most useful in describing the underlying complex process of inter-firm networking while allowing new process dimensions to emerge. Using a process approach (i.e., input-process-outcome), the various dynamic phases involved in developing start-up relationships are examined from the initial encounter to the conclusion of an accelerator program, from which a theoretically conceived framework is derived. The purpose of this framework is to gain deeper theoretical insights into the dynamic nature involved in network building among nascent entrepreneurs and the aspects of entrepreneurial behavior that may influence this process.

1.6. Research findings and contribution

This paper provides an overview of the different phases and types of network formation between early-stage start-ups in the context of an accelerator program. The theoretically derived process model illustrates the social interaction and change mechanisms of network formation between start-ups in a temporal context. Overall, the results of my study point to the importance of dynamic and long-term network building of start-ups that goes beyond their initial phase and highlight the changing importance of different types of relations depending on the respective stage of the entrepreneurial process. They also raise critical questions about the accelerator's current role in promoting network formation and indicate the need to reassess previous program design and management practices.

The theoretical findings and the conceptually derived process model provide several novel insights relevant to academic research, entrepreneurial practice, and individual founders. Firstly, they improve the holistic understanding of the dynamic forms of entrepreneurial network formation based on the changing personal and organizational needs of founders underlying the dynamic nature of entrepreneurial processes. Secondly, they contribute new theoretical insights to social science literature and previous network theory on entrepreneurship by showing that social interaction and network formation among founders is a relevant strategy for start-ups, especially in their early stages. Finally, they reveal how the social environment of start-ups, for example, the environment of an accelerator, may play an essential role in the development of collaborative networks and the strategic management of relationships between founders. Understanding these principles not only improves our understanding of the associated mechanisms within the accelerator environment but also contributes to our overall understanding of the ecosystem's complexity as a whole. It further enhances our understanding of the requirements and opportunities of building and managing an accelerator program in terms of sustainable peer-to-peer engagement. Consequently, the inherent potential of understanding how start-up networks work makes their promotion, support, and development a promising endeavor.

1.7. Structure of the thesis

The remainder of this paper is structured as follows: In the first step, I will provide an overview of theory and previous academic literature combining the established research field of social networks with the still young discipline of entrepreneurship. This part is followed by a description of the research process, providing an overview of the methodology, including the chosen techniques for collecting and analyzing the empirical material. The subsequent section introduces the results of my analysis, followed by an interpretation of the findings in the discussion section. Finally, the study concludes with limitations and possibilities for future research.

2. Theoretical background

The following section reviews the core literature on social network theory. In particular, it highlights areas relevant for a better understanding of the social nature and characteristics underlying entrepreneurial behavior, as well as those influencing innovation capacity.

2.1. Social networks

By understanding how social networks are structured, what elements they consist of, and how they function, it is

[1] I use the terms "inter-firm" and "inter-organizational" interchangeably throughout this paper.

not only possible to design them accordingly but to use them effectively. Furthermore, a better theoretical understanding helps to contextualize the complex relationships between the actors[2] within the entrepreneurial social network of an accelerator and to be able to classify the dynamics underlying the entrepreneurial journey. Since social networks are a topic in their own right and the subject of extensive research and lengthy debates, this section intentionally provides a condensed version of the theoretical underpinnings and broadly accepted concepts that remain within the scope of this paper. In particular, Granovetter's (1973) strength of weak ties theory, Burt's (1992) structural holes theory, and Coleman's (1988) theory of social capital form the starting point of the underlying work providing an approach to the topic of social networks that subsequently allows for further analysis in an entrepreneurial setting.

2.1.1. Definition of social networks

Networks channel the flow of resources, regulate access to information and constitute a significant driver of innovation (Borgatti & Halgin, 2011; Obstfeld, 2005). While some theories (see Burt, 1992) assume that structural holes help actors advance their interests, others (see Granovetter, 1973) emphasize the importance of the strength of connections between networks to obtain opportunity-related information. Although overall research on social networks has increased considerably in recent years, there is still confusion about the fundamental meaning, characteristic elements, and practical applicability underlying network theorizing (Borgatti & Halgin, 2011).

Borgatti and Halgin (2011), therefore, provide a typology regarding the relations between individuals within a network. According to the authors, a social network is characterized by a certain number of nodes (i.e., actors in the network) linked by a defined number of connections of a specific type (i.e., the type of relationship between the actors in the network). These can be distinguished in terms of their content, intensity, and form. Unlike groups, networks do not have natural boundaries and do not necessarily have to be interconnected, allowing multiple unconnected network components to coexist. However, over time, the number of fragmented network components as well as the distance between nodes can change, indicating the dynamic and fluid properties of a network (Lorenzoni & Lipparini, 1999). For instance, two individual actors may initially be characterized by their maximum disconnectedness in the network, although this state may shift or take a different form over time. This would be the case within a friendship or partnership. Essentially, it is assumed that spatial proximity and long-term connections between individuals have a strong impact on the quality of the respective relationship (Lorenzoni & Lipparini, 1999). Moreover, the authors distinguish between two main types of relationships: State-type and event-type ties[3]. States can be defined by their continuity over a specific period of time (e.g., a friend). In contrast, events are countable and transient in nature (e.g., conversations, business transactions, etc.). Both types of connections enable the flow of information, ideas, or goods through interaction between two nodes. Flow frequency, that is, the amount of exchange, can differ significantly depending on the nature and strength of the relationship between two actors (Borgatti & Halgin, 2011).

Beyond these formal and rather abstract definitions, researchers suggest that the overall structure of the relationships determines the possibilities and constraints for the actions of individuals in the network (Borgatti & Halgin, 2011). Studies in this context primarily focus on so-called network structures (i.e., the patterns of relationships between actors) and node positions (i.e., the location or distribution of nodes in the network) and relate them to group and node outcomes. Accordingly, the effect of the structural properties of a network depends on the meaning of the relationships in a particular social context. For this reason, network structures do not have universal but dependent social meanings and consequences (Pachucki & Breiger, 2010). Relational theories are devoted explicitly to the social consequences of structural network properties. In this regard, Granovetter's (1973) strength of weak ties theory and Burt (1992) structural holes theory are two well-known approaches that explain individual advantages such as access to information, bargaining potential, or career opportunities with increasing centrality of actors. In contrast to relational theories, network theories are dedicated to explaining the structural properties of networks from categorical initial conditions. They show that relationships emerge, for example, as a function of spatial proximity, similar social status, or shared organizational affiliation (Borgatti & Halgin, 2011).

2.1.2. Importance of strong and weak ties

A key property of social networks, first addressed by Granovetter's (1973) paper and now a centerpiece of network research, is the strength of relationships. The author developed his interpretation of the degrees of strength of relationships and their usefulness by drawing on findings on processes of relationship formation in social psychology. In doing so, he referred to Heider's (1958) theory on the structural balance of relationships among peers and Davis's (1963) application of this theory to the formation of groups involving individuals with similar interests and interacting partners. The resulting strength of weak ties theory (Granovetter, 1973) has since become an established paradigm and has been increasingly applied as part of the broader theory of social networks (Borgatti & Halgin, 2011). In particular, researchers have focused on the varying degrees of strength of relationships and their impact on the interpretation of network structures or the function of individual elements within a network. In

[2] I use the terms "entrepreneurial actor", "entrepreneur" and "founder" interchangeably throughout this paper.

[3] I use the terms "tie", "connection" and "relationship" interchangeably throughout this paper.

this context, the terminology of strong and weak relationships has been widely adopted by multiple authors in the field when referring to the viability or quality of relationships as part of a social network. Based on Granovetter's (1973) assumptions, the most important factors that influence the strength of ties are the amount of time two people spend together, the degree of intensity and intimacy in terms of the content of time spent together, and the amount of mutual exchange between the actors involved. His research suggests that the stronger the bond between two actors, A and B, the more likely they will share the same social world in which they connect with the same third party, C. In other words, if A knows B and B knows C, chances are high that A and C also know each other. This typification of ties characterizes strongly intertwined networks in which the actors are socially involved with each other, such as within a family or a close circle of friends. Strong emotional bonds of this type enable people to trust others with confidential, private, or otherwise important matters. However, Granovetter (1973) assumes that strongly connected individuals with the same interests and circumstances would frequently share information that is already known and thus redundant. In contrast, the more diverse the personal network between two actors A and D, and the lower the frequency and intensity of their encounters, the more likely they will possess different sets of information and knowledge. He, therefore, assumes that weakly connected individuals can derive more significant benefits from each other by exchanging novel ideas, non-redundant information, or contacts by creating a bridge between their two individual core networks.

This metaphorical bridge between two actors in a social network was further elaborated in Burt's (1992) theory of structural holes. He argued that weak ties only gain significance when serving as bridges for structural holes between multiple networks, allowing the flow of knowledge, information, and value between them. According to the author, innovation takes place at the edges of social networks. In particular, he assumes that people at the edges, near the structural holes, can act as intermediaries or brokers between otherwise unconnected network clusters allowing new opportunities to develop. Overall, their argument differs in that Burt (1992) defines proximity to structural holes, and thus social position in the network, as central, while Granovetter (1973) assumes the strength of ties between individual actors as the key to value creation. Figure 1 and Figure 2 illustrate the different rationale behind the two premises in a simplified fashion.

While both authors emphasize the importance of weak or bridging ties, recent research increasingly advocates the need for strong ties in a network. Krackhardt (1992), for example, agrees with the authors that weak ties are essential for obtaining information but additionally underlines the role of strong ties in exploiting that information as they build trust and function well together.

Although overall, the classification of relationships as strong or weak influences the understanding of network structures and the role of actors within a social network, critics question the practical applicability of the theory, as the notions remain fuzzy in content and therefore allow for different interpretations and applications. In addition to studying social networks in terms of their composition and structures, other streams of research within network analysis examine the quality of the individual connections they are comprised of. Looking at relationships from an instrumental perspective, their value can be measured in terms of social capital, comprising all resources available within the actors' social connections. The following section provides an overview of the concept of social capital as well as an understanding of how it is created and used throughout a social network.

2.2. Social capital

Social capital is integrated into a social structure through interpersonal relationships rooted in social networks (Coleman, 1988). In fact, most individuals are embedded in social contexts and therefore maintain social relationships, the benefits of which they can take advantage of (Kim & Aldrich, 2005). The concept of social capital is arguably one of the fastest-growing research domains in network studies and has significantly increased interest in sociology, political science, and business (see Burt, 1992). Most theories in this area examine how relationships between actors can be used as capital and how the value of these relationships can be determined. Thus, in its broadest sense, social capital can be defined as the value of connections or the sum of resources resulting from relationships with others (Baron & Markman, 2003; Borgatti & Foster, 2003). According to the argument on structural holes, social capital can also be created through a network in which actors (i.e., brokers) can mediate connections between otherwise separate segments (Burt, 2001). Burt (2001) believes that through the function of brokerage, the construct of social capital is given a more precise meaning than through a network of strongly connected actors.

To answer the questions of where social capital comes from, how it can be used, and what criteria to apply to assess its value, network theories often draw on influential terminologies and concepts of James Coleman (1988, 1990), Pierre Bourdieu (1983, 1985) and Robert Putnam (1993, 1995). However, as with other prominent sociological concepts, the term itself has become increasingly vague, and the underlying meaning is elusive due to its extensive application in many fields. To date, a number of definitions have evolved over the years, demonstrating the concept's widespread use within the academic community. For the underlying research topic, however, it seems necessary to establish a consistent understanding to be able to follow the subsequent theoretical considerations. A few of the most established definitions are outlined in the following to illustrate the diversity of perceptions on this topic and to subsequently establish a terminological consensus for the course of this thesis.

2.2.1. Definition of social capital

Coleman (1988), one of the pioneering theorists in the field of social capital, defines the term as a function of social structure creating advantage: "Social capital is defined

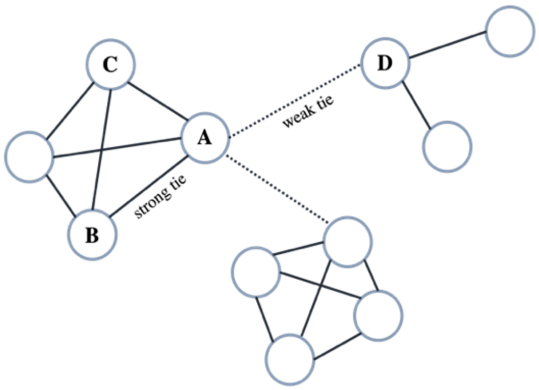


Figure 1: Strong and weak ties in a social network
(Source: Own illustration modified from Borgatti and Halgin (2011, p. 4))

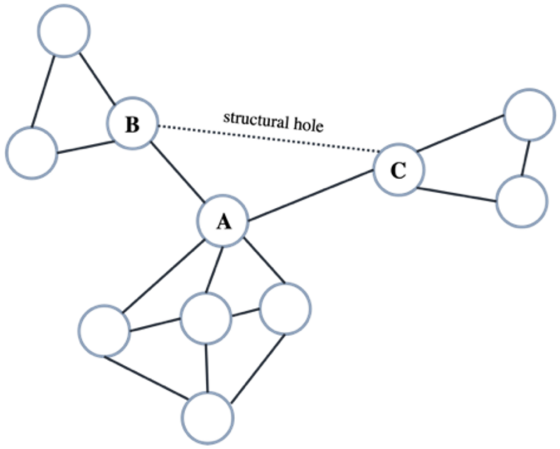


Figure 2: Node A's bridging role in a social network
(Source: Own illustration modified from Borgatti and Halgin (2011, p. 4))

by its function. It is not a single entity but a variety of different entities, with two elements in common: They all consist of some aspect of social structures, and they facilitate certain actions of actors - whether persons or corporate actors - within the structure" (p. 92). Bourdieu's (1985) definition refers less to the individual action but relates the concept of social capital to the social environment within which an actor is located: "Social capital is the aggregate of the actual or potential resources which are linked to possession of a durable network of more or less institutionalized relationships of mutual acquaintance or recognition" (p. 248). Following Putnam's (1993) definition, the concept of social capital usually includes norms and values that arise within a relationship structure and have a positive effect on people's solidary action: "Social Capital here refers to features of social organizations, such as trust, norms, and networks, that can improve the efficiency of society by facilitating coordinated action" (p. 167).

Even though the perspectives cited above have different definitions and approaches, they agree on the generation of social capital. The available social capital is derived from individual social relationships that create advantages for the individual or group. Accordingly, social capital can be broken down to the general idea that some people enjoy advantages because of their relationships with others (Burt, 2004; Lin, 2001).

2.2.2. Characteristics and forms of social capital

Development of Social Capital

When studying the origins and formation of social capital, sociologists tend to follow a bottom-up approach (Coleman, 1988; Portes, 1998), while it is predominantly political scientists who take the contrary view of the top-down approach (cf. Evans, 1996; Levi, 1996; Woolcock, 1998). According to the bottom-up perspective, social capital is primarily the result of individual social relationships. By cultivating these relationships, social capital can be aggregated into a collective resource creating shared benefits (Burt, 1992; Portes, 1998). This approach assumes that networks are formed through cooperative action, which constitutes the basis for social behavior and generalized trust. Especially trust between people is an essential element in social relationships and affects the

extent to which social capital is maintained or decays (Putnam, 1995; Yen et al., 2015). In contrast to the bottom-up approach, the top-down perspective views social capital as created by state or public institutions promoting standards and values through regulation and fostering trust in the wider community (Levi, 1996).

Both approaches are based on the idea that the more social capital a group has, the stronger its cohesion. However, the controversial question discussed in both approaches is whether there is a predominant direction from which social capital is generated. Thus, the question remains whether a state's or organization's respective governance leads to mutual trust within social networks or whether the individual actors and the relationships at the micro-level drive the emergence of social capital (Fuchs, 2020). Especially in network analyses, it, therefore, seems crucial to account for the context in which social capital is being studied and to consider the corresponding level of investigation (micro-, meso- or macro-level).

Forms of Social Capital

Another dimension based on which social capital can be distinguished is the type of connection between individuals or groups. Putnam (2000) differentiate between bridging and bonding social capital. The former can be attributed to Burt's (1994) notion of structural holes between networks, describing relationships with external people outside the individual's core network (Adler & Kwon, 2002). This type of bond is typically reflected in heterogeneous group compositions in which members differ, for example, in terms of status, generation, or gender. On the other hand, bonding social capital is based on reciprocity and norms of trust. It refers to cooperative behavior or support measures toward people similar to oneself, often acquaintances, relatives, or friends. As the names already reveal, bonding social capital, in contrast to bridging social capital, has an effect on an actor's internal or existing network. It can, therefore, also strengthen identity or cohesion within the network based on shared norms and a mutual sense of trustworthiness (Coleman, 1988). Bonding social capital is what Coleman (1988) defines as closure within a network.

While the two forms of bridging and bonding social capital can be separated in theory, they are not necessarily mutually exclusive in practice (Fuchs, 2020). In particular, both views can provide significant value depending on the context and the objectives of the relevant people involved (Adler & Kwon, 2002). On this basis, the theoretical distinction allows us to examine the characteristics of social capital from multiple perspectives in different contexts.

Value of Social Capital

In network research, social capital is often ascribed an inherent benefit. It can be used as a resource to maximize utility underlying the economic principle of rational action and expressed in support or assistance services between two or more individuals (Coleman, 1988). Coleman (1988) similarly assumes that with the help of interpersonal relationships and networks, goals can be achieved that would otherwise be unattainable. The resources resulting from social capital may be another person's knowledge, social or financial support, or other forms of assistance (Putnam, 2000). In addition, social connections between actors within or between networks can facilitate the acquisition of valuable information, which is also an incentive for cooperative action. This form of social capital is consistent with the findings of the theories of Granovetter (1973), who places information sharing at the center of his approach. However, according to Burt (1997), such information advantages do not necessarily arise through an increased number of network connections but through strategic positioning within the network. In particular, heterogeneous relationships can help gain access to more varied information. Beneficial opportunities, such as referrals, often accompany these benefits as individuals with a diverse portfolio of contacts are often more attractive to external parties. Additionally, individuals with networks rich in structural holes benefit from the complementary properties of social capital. Compared to individuals with a relatively homogeneous or limited network, these people are more likely to identify and exploit rewarding opportunities, further leveraging their individual capabilities, usually referred to as human capital (Burt, 1997).

The abstract and versatile nature revolving around the concept of social capital is nothing new. What is new about it, however, is that established theories of social capital are increasingly leaving their traditional areas of application and moving into new areas of sociology, including start-up research. Still, the literature on entrepreneurial networks currently seems to hold different perspectives regarding the underlying principles of social capital and the importance of different types of networks throughout the entrepreneurial process (Scott, 2011). The following section, therefore, aims to improve the understanding of the role of social networks and collaborative relationships in the context of entrepreneurial activities.

2.3. Social network theory and entrepreneurship

The success of start-ups depends, among other things, on the ecosystem surrounding them, that is, the relationships and interactions between actors that shape their immediate social network (Greve & Salaff, 2003). The widespread image of the entrepreneur as an independent and autonomous leader is long outdated. Entrepreneurship is a social phenomenon embedded in a social context. Indeed, following premises from social network theory, entrepreneurship is a phenomenon embedded in networks of enduring social relationships (Walker et al., 1997). The inclusion of actors in networks of different relational content has recently prompted a large body of research that has shifted the focus from examining the individual characteristics of entrepreneurs to understanding the relationships between them. Most researchers agree that the individual entrepreneur should no longer be viewed in isolation but as a social entity enclosed

by a broader social environment, the effects of which influence the entrepreneur's behavior and actions (Aldrich & Zimmer, 1986).

2.3.1. Entrepreneurial networks and networking

In the context of entrepreneurship research, network theorizing has received wide recognition and has long been argued as one of the essential factors for acquiring resources, providing emotional and professional support, reducing the risk of failure as well as improving innovation performance and competitiveness (Baum et al., 2000; E. L. Hansen, 1995; Jack et al., 2008; Pittaway et al., 2004). Theories of alliance networks, in particular, assume that the formation of alliances, especially among young and resource-constrained firms, will help them overcome liabilities of newness and smallness and increase their overall chances of survival (Gulati, 1998; Teece, 1992). Accordingly, entrepreneurs organize and coordinate available resources in the social network by interacting with others to pursue or exploit an entrepreneurial opportunity (Baum et al., 2000).

To date, most studies of entrepreneurship and entrepreneurial behavior from a social network perspective are concerned with investigating and understanding different types of relations between actors in the network who provide the resources and knowledge necessary for starting a business (Johannisson, 1988; Larson, 1991). In this regard, Aldrich and Zimmer (1986) suggest that the strength of these types of relationships primarily depends on the frequency and reciprocity of the relationship (i.e., the expectation of the favor being returned). According to the authors, a network of strong relationships helps the entrepreneur to activate cognitive and emotional resources such as self-confidence and to remain motivated throughout the entrepreneurial journey (Aldrich & Zimmer, 1986). Due to high information redundancy, however, networks whose actors are highly interconnected are assumed to have lower innovation capacity. For this reason, entrepreneurs do not necessarily seek the support of those closest to them but instead form weak ties with actors they believe will rationally benefit their business. Although weak relationships are less reliable, they are a particularly important source of relevant and diverse information for the entrepreneur (Aldrich & Zimmer, 1986; Granovetter, 1973). Rost (2011) likewise examines the strength of ties in the entrepreneurial environment. She argues that combining strong and weak ties is crucial for enhancing knowledge transfer and innovation capacity. According to the author, weak ties facilitate access to peripheral network positions and thus to new knowledge and ideas, while strong ties ensure the translation of these ideas into innovative solutions, indicating the complementary potential of both conditions. Overall, a personal network structure with a balanced mix of strong connections within the core cluster and a large number of weak ties that form bridges to other network clusters has been found to be particularly conducive to the founder's economic success (Rost, 2011; Uzzi, 1997, 1999).

2.3.2. Entrepreneurial network dynamics

Aside from the difficulty of studying the different structures of start-up relationships and classifying them adequately, a relatively recent line of research focuses on examining how different types of relationship networks change over time (Jack et al., 2008). As Burt (1982) noted, in addition to examining the structural properties of networks, network analyses should also consider associated procedural changes. Understanding how relationship structures change over time can provide insights not only into how relationships are formed but also into how they affect subsequent relationship formation. Typically, two fundamentally different perspectives are considered in this context. The former identifies network structures based on the development of the entrepreneurial project. The opposite view attempts to describe the types of relationship structures that impact the development of the venture (Lamine et al., 2015). While some researchers argue that relationship structures may change over time due to external influences, others assume that entrepreneurs consciously adapt their network structures based on strategic decisions (Jack et al., 2008; Stuart & Sorenson, 2007). At the same time, however, critics in this context point to the limited applicability of rational network design due to the uncertainty underlying the entrepreneurial processes (Engel et al., 2017). In this context, for example, Nebus (2006) posits a heuristic theory of network formation arguing that in information-poor or uncertain situations, new connections must be made before their potential value can be assessed.

Although ideas for efficiently building and managing relationships improve our general understanding of how network structures work, individuals' networks often lack the necessary efficiency. Building and maintaining meaningful relationships requires competencies not necessarily derivable from social behavior.

2.4. Start-up accelerators as social networks

Creating an enabling environment for new and emerging companies to overcome some of these challenges is precisely the approach relevant to the accelerator phenomenon. Accelerators play an essential role in this context facilitating access to valuable network contacts and thus shortening the path to appropriate resources. Unlike the established research and literature on social networks, accelerators are a relatively new form of entrepreneurial support organization and are still an insufficiently researched field in which terms and definitions are constantly changing (Goswami et al., 2018). For this reason, it is necessary to only address the critical findings of network theories in the context of accelerator research and to confine myself to one accepted definition of an accelerator to guide the remainder of this thesis. The underlying section, therefore, only presents the broader understanding that provides the necessary framework for examining more specific questions in the context of one particular accelerator.

2.4.1. Definition of accelerators

Start-up accelerators are a relatively novel but rapidly spreading phenomenon within the entrepreneurial landscape, helping prospective founders recognize and navigate the business challenges faced in their early stages of growth (Cohen, Fehder, et al., 2019). Proliferating over the last decade, they have become synonymous with any form of initiative designed to help nascent entrepreneurs compensate for their lack of knowledge, financial resources, and contacts with relevant partners within a protected environment (Cassar, 2004; Shepherd et al., 2000). Because of the novelty of the phenomenon, a considerable amount of recent research has been devoted to answering relevant questions about what accelerators are, what they do, and to what extent they effectively deliver on their core message of accelerating businesses (see Cohen, Bingham, and Hallen, 2019; Cohen, Fehder, et al., 2019). Inside the complexity of answering these questions, a widely accepted definition has emerged identifying accelerators as fixed-term, cohort-based initiatives, providing mentorship and educational components, concluding in a public pitch event, commonly referred to as demo-day (Cohen, Fehder, et al., 2019). In this sense, the accelerator differs from other entrepreneurial support institutions, such as incubators, particularly in terms of its finite duration and cohort-based structure. Beyond that, however, accelerators exhibit some common characteristics with those of incubators. For example, the provision of networks is considered an essential element of entrepreneurial support in both incubator and accelerator environments (M. T. Hansen et al., 2000; Soetanto & Jack, 2013). As with most incubators, participants in accelerators receive access to the accelerator's broader external network, such as universities, companies, and investors. Although access to external networks remains crucial for entrepreneurs to source potentially relevant contacts, recent studies have increasingly turned their attention to investigating internal networks, that is, networks between the founders operating in an accelerator environment (Bøllingtoft, 2012; Krishnan et al., 2020; Soetanto & Jack, 2013). Particularly given the fact that many accelerators co-locate their members under one roof for the duration of the program, the associated emergence of internal networks not only seems obvious but deserves closer examination in this regard.

2.4.2. Networking in accelerators

The success of start-ups essentially depends on the ecosystem surrounding them, that is, the nature of interactions between the actors as part of the local environment and the network connections they create. Making network connections is typically viewed as an essential competency of accelerators (Hallen et al., 2020). Compared to the ubiquitous role of accelerators in connecting founders to the regional ecosystem, the role of the accelerator also includes its ability to foster internal network connections among founding teams (Soetanto & Jack, 2013). There is hardly a mission statement of an accelerator that does not emphasize the importance of social collaboration with peers experiencing similar challenges (Krishnan et al., 2020). While this characteristic may hold for different types of start-up support institutions, social networking seems to be even more prevalent in accelerators due to the intensity of the program, the short duration, the cohort-based structure, and the inherent nature of accelerators to encourage their members to interact and support each other. On this premise, accelerators create collaborative environments for nascent founders, often in the form of an enclosed spatial setting designed to promote physical proximity and opportunity for interaction (Saxenian, 1994). In this context, peer connections between founders constitute a key source of social exchange in terms of advice, professional guidance, and emotional support (Hallen et al., 2020; Huang & Knight, 2017; Saxenian, 1994). Other authors increasingly point to the importance of peer effects in the context of learning from each other and sharing mutual experiences (Kacperczyk, 2013; Nanda & Sørensen, 2010). Ahmad and Ingle (2011) even consider internal social networks between founders as the most essential element of entrepreneurial support organizations. The exact nature of the network and the degree of interaction among founders, however, depends largely on the composition of the cohort in terms of its relative similarity, prior experience, and technical expertise as well as the accelerator's overall organizational and structural design (Cohen, Fehder, et al., 2019).

2.5. Research questions

Despite common academic agreement on the inherently collaborative nature of the accelerator environment, little is known about the respective mechanisms of collaboration and the interactive behavior of accelerator tenants: How do start-ups connect? How do networks of founders form? And what role does the accelerator play in channeling collective experiences among its tenants? In this study, I use a process-oriented qualitative approach (Denzin & Lincoln, 2005) to develop an inductive theoretical model (Strauss & Corbin, 1990) seeking to shed light on the various interaction dynamics among founders. The accelerator context seems to provide an excellent ground for studying social interaction mechanisms, internal relationships, and the emergence of founder networks. In particular, locating the accelerator at the meso-level of analysis allows for an in-depth assessment of its mediating role between individual actors and the broader social network of founders. In this regard, I seek to answer the following research questions:

> *(1) How do early-stage start-ups form, use, and develop relationships with other founders?*
>
> *(2) How does the importance of different relationships change in light of the dynamic nature of entrepreneurial activity?*
>
> *(3) How does the involvement of accelerators impact the process of relationship building between founders?*

3. Research methodology

I adopted a qualitative inductive approach embedded in a single case study to gain a deeper understanding of the underlying social mechanisms related to the accelerator landscape and to formulate new theoretical foundations on the dynamics of the entrepreneurial process.

3.1. Research context and design

Compared to the amount of quantitative variance-based theories providing answers to the *what*, entrepreneurship research often fails to account for the underlying temporal nature of the phenomena under study (Gartner & Birley, 2002; Van de Ven, 1992). In addition, phenomena in entrepreneurship, including concepts such as social capital or entrepreneurial activity, often exhibit unusual characteristics or substantial variations in the data (Crawford et al., 2015; Davidsson, 2016). In such cases, quantitative research approaches often involve cumbersome modeling procedures to adjust for biases and outliers in the data. In contrast, qualitative research is less restrictive and allows for a more detailed examination and interpretation of deviant findings or extraordinary outcomes (Aguinis et al., 2013). Therefore, to advance the study of research and to fully unravel the complexity of entrepreneurial dynamics, researchers are increasingly advocating the use of process-oriented qualitative methods to capture processes and sequences of events holistically and over time (Shepherd et al., 2015; Van de Ven & Engleman, 2004). Due to the lack of in-depth understanding of network development processes in entrepreneurial research, I employ a qualitative, exploratory single, and holistic case study approach. Qualitative data collection proves to be a particularly suitable method for studying complex processes, as it allows phenomena to be holistically captured, taking into account their development over time (Langley, 1999). In particular, choosing a single case is justified when the relevant case provides access to a social field that is still relatively young in nature and/or lacks theoretical understanding (Eisenhardt, 1989). In addition, case-based data offers explanations for the *how* and *why* of the phenomenon under study as they guide the underlying work (Eisenhardt, 1989; Eisenhardt & Graebner, 2007). These explanations often form the basis for the emergence of theoretical constructs that highlight the deeper narrative description of the process under study expressed in an underlying pattern of events (McMullen & Dimov, 2013; Miles & Huberman, 1984; Pentland, 1999).

3.2. Sample selection and data collection

Given my interest in exploring and understanding the dynamic phenomenon of networking among start-ups at the micro level, I adopted a purposive sampling approach. This approach seemed most promising for obtaining rich data about the phenomenon under study and a high degree of explanatory power across all individual process steps associated with it (Patton, 2002; Yin, 2009). To initiate my research, it is essential to identify an appropriate research context that would (1) allow a transparent and in-depth view of the underlying dynamics in a real-life setting and (2) provide access to both potential sample candidates as well as other relevant stakeholders involved in the formation process (Yin, 2003).

Start-up accelerators provide the natural context that satisfy both criteria. The particular setting and dynamic nature of an accelerator offer various insights into the characteristics of start-up relationship formation: How relationships are formed, used, and sustained. Studying the accelerator environment in this context not only helps to understand certain determinants such as individual characteristics and cognitive processes involved in relationship building among founders but also provides the basis for systematically managing both contextual and sociological conditions, such as information and resources availability, that influence the formation process (Dimov, 2011). Additionally, the fact that start-ups typically do not know each other personally prior to joining an accelerator program increases the likelihood of natural network effects being observed (Cohen, Bingham, & Hallen, 2019). Finally, by providing a comprehensive and thorough account of the entrepreneurial experience in a real-world setting, observable principles and interrelationships among internal actors also suggest transferability to other contexts (Gioia et al., 2013).

I initiated my research by investigating an accelerator of a large technical university in a German metropolitan region. The accelerator provides access to a mature institutional ecosystem that supports young entrepreneurs in implementing and scaling up technology-driven business ideas. In particular, the accelerator focuses on high-tech start-up teams from diverse industries in their early formation phase. Key support areas include customer acquisition, business model development, and venture capital funding. As part of a larger entrepreneurial network, the accelerator serves as an interface for start-ups to connect with investors, business angels, mentors, industry partners, and innovative companies. The program includes a 12-week curriculum during which founding teams benefit from various coaching, mentoring, and workshop formats. During this time, they also have the opportunity to use the accelerator's services and office space, with the option of a three-month extension of use upon completion of the program. At the time of the study, the accelerator accommodated a total of 14 early-stage (i.e., pre-seed or seed stage) technology-based start-ups, half of which participated in the program on-site and half remotely, attending events or workshops only on a bi-weekly basis. All participants identified themselves as founders or co-founders of the company while none of the respondents reported any previous exit through the sale of an earlier venture ensuring homogeneity of teams in terms of their entrepreneurial experience. I contacted all 14 start-ups, 10 of which were available for the complete duration of the study and willing to communicate their experiences and attitudes transparently and reflectively, which I considered necessary for the formation of a reliable theoretical framework (Bernard, 2017).

To best capture participants' voices and give theoretically scientific meaning to their experiences, I conducted two rounds of semi-structured interviews with one representative from each start-up, which serve as the primary source of data for my study (Gioia et al., 2013). Interviewing participants twice is consistent with my goal of exploring the dynamic process and development behind the phenomenon of relationship formation. The first round of interviews was conducted three weeks after the program began to ensure a certain level of familiarity among the participants, and another round at the end of the program to capture possible changes in relationships among founders. Studying network formation in real time helps me overcome the methodological challenges of hindsight or recall bias common in retrospective studies (Davidsson & Honig, 2003). An overview of the relevant start-ups can be found in Table 1.

As the qualitative single case study is not a method as such but rather a procedure in which different methods can be combined, I collected data from three types of sources, all in the form of semi-structured interviews: (1) Start-up representatives from the active cohort at the time of the study, (2) start-up representatives from the previous cohort, as well as (3) representatives from the accelerator team. Interviewing previous start-ups allows me to assess the validity of the research results by comparing retrospective and real-time accounts while serving as valuable input for the guiding questions addressed in the second round of interviews. The five representatives of the accelerator team were chosen as key informants based on their extensive insights into relationship building among start-ups and/or familiarity with internal strategies and policies around network building within the organization. Informants were recommended by program management to ensure their eligibility and competence in answering my questions about the topic under study. This led to the identification of two venture consultants (later referred to as coaches), two external advisors (later referred to as mentors), one of whom was the former head of operations, as well as the managing partner of the program. By incorporating multiple data sources from different individuals into the data collection process, additional perspectives can be added to the empirical foundation of the findings from participant interviews to account for a nuanced understanding of the phenomenon of interest (Neergaard & Ulhøi, 2007; Yin, 2009). This type of data source triangulation allows for validating the data by minimizing the possibility of data bias due to falsified or distorted responses from interviewees (Kumar et al., 1993). Appendix 1 and Appendix 2 provide an overview of the start-ups from the previous cohort and accelerator team members, including a summary description of their roles and associated tasks within the program.

In total, I conducted 24 semi-structured interviews (including both interview rounds), each of which lasted 20-90 minutes (excluding introductory conversations), was digitally recorded, and transcribed to ensure the completeness and accuracy of the data. Except for one interview, which took place in person, all interviews were conducted via the video communication tool MicrosoftTeams or GoogleMeets, with permission to record the interviews obtained in advance. The statements of the interviews conducted in German were translated into English. To create a common ground on which interpretive comparisons could be made, I followed a similar set of questions across all interview groups essential to answering the research questions while leaving enough room and openness for new perspectives and concepts to emerge. Therefore, I revised the initial interview protocol several times to adjust specific questions, sharpen the focus in light of my research question, or account for possible twists and turns in the process (Glaser & Strauss, 1967). Appendix 3 to Appendix 5 include the general interview protocols for the three groups of respondents, with adjustments to the initial questions indicated in italics. The individual questions of the first round of interviews were divided into two main categories addressing (1) the start-ups' perception regarding the internal network within the accelerator and (2) the role of the accelerator in forming and developing relationships between the start-ups. Follow-up questions were asked in the second round of interviews to capture any changes in the relationships between the start-ups at the beginning and end of the program. Constant comparison allowed me to iteratively compare the data across informants and over time (Strauss & Corbin, 1990). To ensure anonymity, the real names of the respondents were replaced by consecutive digits.

3.3. Data analysis

Following a grounded theory approach allows theoretical models to be derived from the available data in order to explain social phenomena in their natural environment (Strauss & Corbin, 1998). Additionally, it is one of the few methods that specify not only how data should be collected, but also how it should be analyzed, which directly involves the researcher's inner attitude toward the subject. In this context, it is particularly important to separate the data from existing cognitions and attitudes (Suddaby, 2006). Therefore, in order to analyze the underlying process in an unbiased and open-ended manner and to provide sufficient explanatory theoretical insight, I deliberately refrained from a quantitative research approach at this point. Instead, I applied an inductive coding approach by strictly following the underlying data throughout the process of data analysis to allow theoretical constructs to emerge (Glaser & Strauss, 2008). Moreover, an inductive approach is particularly appropriate when the specific context can provide additional important insights into the studied phenomenon (Saunders et al., 2012). Previous theories were only considered after the evaluation process was completed in order to allow for the comparison of my own findings with those of the literature. Since the underlying accelerator involves a larger social unit, it was important to compare patterns of interpretation and action of the individual founders in order to develop propositions about the respective social unit and the patterns of action and selection decisions typical for it. Superordinate patterns and constructs within the social network emerged from the com-

Table 1: List of interview partners - start-up sample

(Source: Own illustration)

Interview Partner	Role	Venture Stage	Technology	Industry	Type of Participation
Start-up 1 (S1)	Co-Founder, CFA	pre-seed	BioTech	Diverse	remote
Start-up 2 (S2)	Co-Founder, CEO	pre-seed/ close to seed	AI, NeuroTech	Gaming	on-site
Start-up 3 (S3)	Co-Founder, CEO	pre-seed	AI, DeepTech	Inventory, Sustainable Production	on-site
Start-up 4 (S4)	Co-Founder, CEO	pre-seed	Robotics	Mobility	on-site
Start-up 5(S5)	Co-Founder, CEO	pre-seed/ close to seed	LogTech (Logistics), SaaS	Transportation, Supply Chain	on-site
Start-up 6 (S6)	Co-Founder	pre-seed	HealthTech	Healthcare	remote
Start-up 7 (S7)	Co-Founder	pre-seed	FoodTech	Food	remote
Start-up 8 (S8)	Co-Founder	pre-seed	AgTech (Agriculture)	Agriculture	remote
Start-up 9 (S9)	Co-Founder, CTO	pre-seed	LegalTech	Diverse	on-site
Start-up 10 (S10)	Co-Founder	pre-seed/ close to seed	UrbanTech	Construction	remote

parison of roles and action patterns of the individual members.

I began my data analysis with an open coding approach by assigning codes to the collected data in order to break down the overall phenomenon into sub-processes and to develop possible explanatory building blocks for the evolving theoretical framework (Strauss & Corbin, 1998). Categories were created based on the in vivo coding approach adhering to the terms and language used by the informants (Strauss & Corbin, 1998). I used descriptive expressions when an in vivo code did not seem useful or was not available. These resulting first-order codes allowed me to create narratives while aligning the data as closely as possible to respondents' perspectives (Langley & Abdallah, 2011). In the next step, I applied axial coding (Strauss & Corbin, 1998) to search for possible connections between the developed categories, finally grouping them into so-called second-order themes at a higher level of abstraction (Gioia et al., 2013). By further comparing the data, I was eventually able to reduce the data to a minimum of seven aggregate dimensions which form the basis for the theoretical data structure. The graphical representation of the data not only allows to understand the elaborated logic of the data but also to visually recreate the process of analysis - a crucial element in demonstrating qualitative rigor in research findings (Pratt, 2008; Tracy, 2010).

I coded the data using MAXQDA, a qualitative data analysis program. This allowed me to structure the data and group related concepts into categories. Throughout the coding process, I followed an iterative approach, repeatedly reviewing and revising the codes and categories that emerged from the data to identify potential distinguishing features and/or commonalities (Strauss & Corbin, 1998). I followed this process until theoretical saturation was reached and no more new substantive insights could be gained through further interviews (Glaser & Strauss, 1967). The resulting coding scheme comprised 55 summarized first-order codes derived from the individual respondents' statements. I clustered the first-order codes into second-order themes until no more new themes could be identified. This resulted in a total of 21 second-order themes that were evident across all interviews. Finally, the themes were grouped into seven overarching aggregate dimensions. Only findings that could be confirmed by multiple responses from different interview partners are provided in the subsequent quotes. The overarching data structure derived from the data collection and analysis is depicted in Figure 3.

The following section presents the main findings that emerged from the extensive coding and analysis of my data. I then interpret the results to derive a central theoretical construct as a basis for practical application and subsequent research.

4. Results

Before delving into the detailed presentation of the results of my data, I would like to point out that the interviewed start-up teams had already been actively participating in the accelerator program for three weeks prior to the start of the first round of interviews. This was an essential prerequisite for the initial data collection to assume a certain level of interaction between the teams. In the case of my sample, the majority of all start-up teams were at an early stage of the development process (pre-seed or between pre-seed and seed stage) with a professional network that was not yet well established. This fact, coupled with the teams' full-time participation in the program, fostered an environment that allowed the teams to initiate networking activities and build relationships with their peers. All interviewed

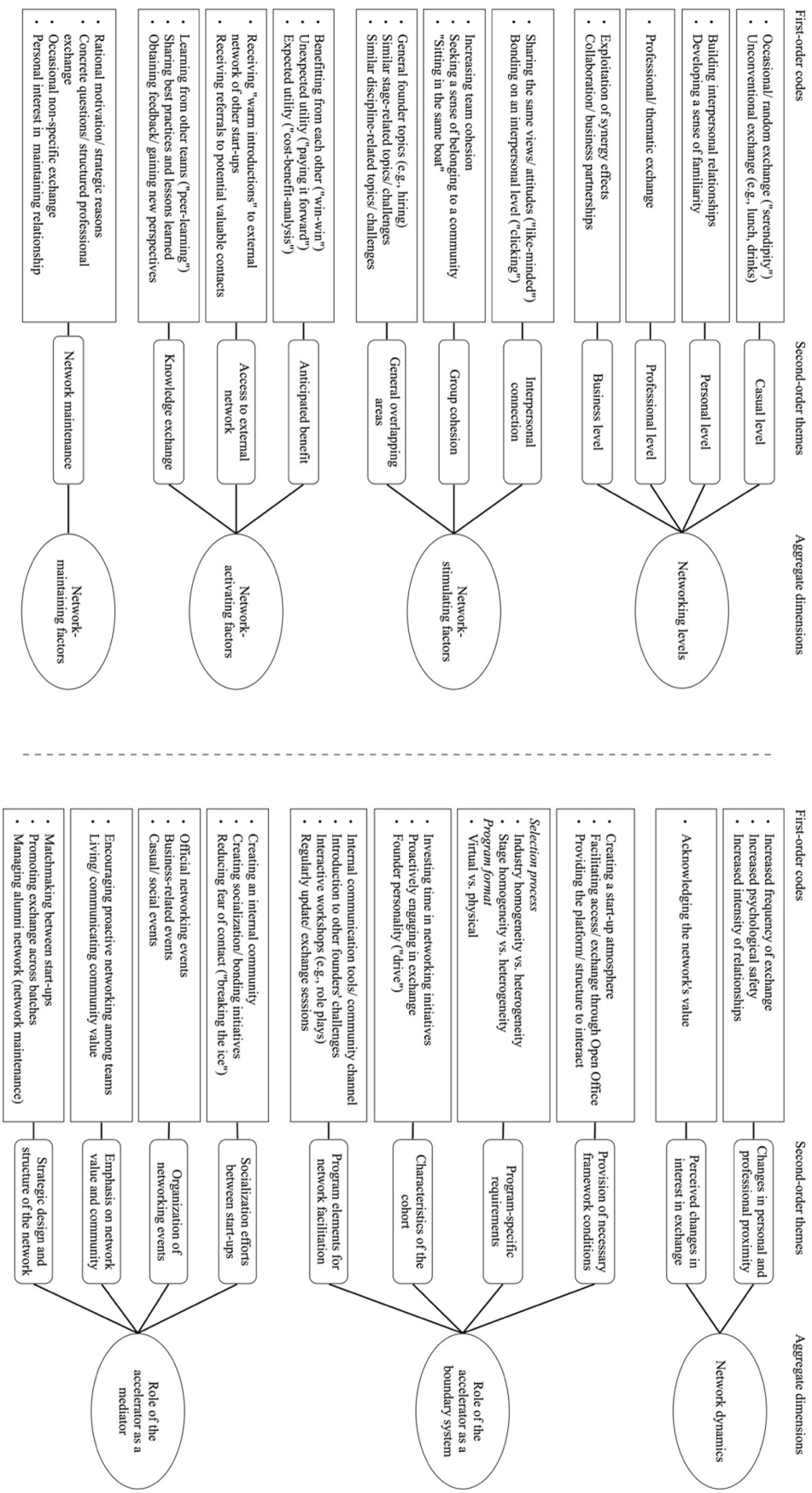


Figure 3: Data structure (Source: Own illustration)

start-ups confirmed that the individually attributed value of networking with other founders increased over the course of the program.

But how did networking efforts evolve from the teams' initial encounters to later points in the program? What individual attitudes underpinned the various forms of interaction and relationships? And what role did the accelerator team play in this context? In analyzing the dataset, distinctive patterns emerged that I considered relevant to understanding the phenomenon under study and answering the research questions. The relevant patterns were grouped into the following seven overarching themes: (1) *Networking levels*, (2) *network-stimulation efforts*, (3) *network-activating efforts*, (4) *network-maintaining efforts*, (5) *network dynamics*, (6) *role of the accelerator as a boundary system* and (7) *role of the accelerator as a mediator*. The respective patterns identified within the dimensions are presented in the following sections, along with an overview of illustrative statements that underpin the findings (Appendix 6 to 26). Subsequently, propositions and a process model are derived to allow meaningful contextualization of the research findings and to better understand the dynamics of network formation over time. To preserve anonymity, I replaced the program's official name with the fictitious name "X Combinator" in the following sections.

4.1. Networking levels

When analyzing the founders' statements about their perceived interaction mechanisms with other founders both during and after the program completion, I identified four different networking types that could be classified into the following dimensions: (1) *Casual*, (2) *personal*, (3) *professional*, and (4) *business level*.

Casual level

The *casual level* of networking refers to loose, unconventional interactions between teams or individual founders. Exchanges of this type mostly occur randomly and unplanned. The casual nature of this type of networking is reflected in the form of interaction, which is limited primarily to conversations before or after workshops, business meetings, or other program-related activities, as one of the founders put it: "It's just being in the [office] and you randomly see someone and it's like, 'oh, you want to grab lunch or how's it going?' [...] You have these bit more random conversations, which are also really nice" (S7). Irregular interaction could indicate a negative impact on the intensity of the relationship. However, when asked about the impact of this type of relationship, most founders did not confirm this assumption. For example, one founder reported an unconventional exchange with another team at the coffee machine, after which one of the team members shared personal contacts of the team's investors. Such incidents suggest that it is not necessarily the nature of the relationship that influences information sharing among teams. Overall, it appeared that, in fact, most of the relationships could be classified as informal, casual relationships that helped founders "break down barriers" (S4) and "facilitate communication after" (S4). A key difference I found in the data was that on-site founders, in particular, mentioned this type of relationship, as opposed to virtual teams, which experienced less daily interaction with teams due to geographic distance.

Personal level

As the name suggests, the *personal level* of networking refers to a type of connection that is primarily interpersonal in nature and closely linked to building trust and mutual emotional support. During the interviews, I noted that a personal relationship was less due to the start-ups' focus or thematic overlap but a result of personal interest or individual attraction to certain character types: "In the end, you make friends with people or not. I don't think it necessarily had to do a lot with the start-up they were doing" (S7). Overall, data revealed that, in particular, relationships that included both overlap in content and a personal connection was found to be most valuable for teams to support each other throughout the program.

When asked at what point personal relationships developed, informants agreed that the initial opening event essentially helped to build an emotional connection and establish a certain level of trust between the founders. The benefits of a personal relationship were particularly evident on a psychological level, helping founders "deal with stress and pressure" (O-M) throughout their entrepreneurial journey. Finally, a clear positive relation between personal connection among the teams and the longevity of the relationships beyond the program was found. In contrast, virtual teams developed fewer personal relationships relative to on-site teams. As already mentioned above, this may be due to the fact that virtually participating teams had fewer opportunities to attend in-person events and face-to-face interactions. Physical proximity thus seems to constitute an essential element for the development of personal bonding among founders.

Professional level

The *professional level* of networking forms the third category of network relationships and summarizes the professional-thematic interactions between teams or exchanges in the form of business-relevant information. In particular, program formats such as workshops or expert talks, which were thematically based on the given curriculum, served as a starting point for further content-related discourse among the teams: "We had different modules, fundraising, sales, business development, team building, and product [...] we structured the program systematically or strategically so that people then [...] exchanged ideas about a particular topic" (O-M). The bi-weekly update sessions between the teams to share progress and challenges proved particularly useful for sharing specific thematic issues with the teams. Founders noted that the thematic exchange was beneficial to get "different impressions" (S7) on open

questions of general topics such as "financing strategy" (S7), "accounting" (S2) or "hiring" (S9). Based on the interview statements, I observed that the thematic exchanges on a professional level primarily took place with founders who had "similar topics" (S1) or "overlaps" (S1) in terms of content.

Business level

> "There were two start-ups doing certain manufacturing of components. One of them was doing subtractive manufacturing, and the other one was doing additive manufacturing. And it happened that the additive manufacturing guys needed a piece of a certain material, they couldn't do with additive manufacturing. So they got it from the subtractive manufacturing, guys. So they both work together to please an external third party client. So that's, for example, another type of business relationship." (C1)

As seen from the above example, the *business level* of networking describes a business connection between start-ups. Business relationships within the accelerator program involved collaborations or partnerships between teams based on complementary skills, but also in "cross-hiring activities" (C2), when one team hired an employee from another team, which usually happened when one team did not perform as expected. One of the founders reported a possible collaboration with another team after the program: "It could be that we also do a project with [S5] after the program because it fits very well into our strategy. Without [X Combinator], I don't think the contact would have ever happened" (S3). Another founder mentioned considering teaming up with an alumni founding team to offer a "more differentiated value proposition in the marketplace" (S6). In addition, there were also instances where a start-up offered its product or service to another team, which was a common occurrence throughout the program, as reported by one of the coaches. Although business-level interactions accounted for the least amount of networking activity, founders overall confirmed that they could generally imagine potential partnerships with particular start-ups. Illustrative quotes of the different types of *networking levels* are summarized in Appendix 6, 7, 8, and 9.

4.2. Network-stimulating factors

After identifying four main categories of networking levels, I looked for emerging patterns in the different stages of relationship formation to assess potential changes in the course of the program.

Network-stimulating factors emerged as the first pattern in the data, describing the initiators or motivational reasons on the part of the start-ups for reaching out to other founders throughout the program. The main reasons for initiating contact can be classified into the following five groups: (1) *Interpersonal connection*, (2) *group cohesion,* and (3) *general overlapping areas*.

Interpersonal connection

Interpersonal connection defines an incentive to connect with certain start-ups based on interpersonal motives and shared beliefs. This type of initiator is evident mostly at the personal network level: "There are a few people that you just click [. . .] I think that's natural. I mean, same with you, you have a few friends, and those are people you click, it's nothing against the other people, it's just you simply click" (S5). Especially at the beginning of the program, initial relationships were built based on interpersonal connections rather than thematic overlap or personal interest in the other start-ups' businesses. Although personal relationships with start-ups were established primarily at the beginning of the program, some founders reported that as the frequency of contact with other start-ups increased, personal relationships continued to develop weeks later. This could be related to the fact that the start-ups communicated more openly with each other as the program progressed, allowing for similarities to be discovered among the founders over time.

Furthermore, one founder indicated that the perceived level of mutual sympathy and the interpersonal bond between teams also seemed to impact their willingness to collaborate throughout the program. In addition, private matters were more likely to be shared with teams with whom there was a personal relationship: "With some, we basically have a very friendly relationship because they're funny brands with whom you want to spend time together. [. . .] With them you feel more free to discuss private things, like, P&L and stuff like that" (S5). This was confirmed by another founder, who associated relationships based on interpersonal attraction with higher levels of trust than those he experienced with less closely connected teams.

Group cohesion

The start-ups' initiative to network with other founders within the accelerator was also related to their desire to feel part of a group of like-minded people, as one of the founders described:

> "One of the benefits of being part of the program is that you're actually part of the batch [. . .] you are together with like-minded people, you know, your friends do not understand, your family does not understand what you go through as a founder. People who are part of the batch do because they're doing exactly the same thing. So I think it's super valuable, the networking, and there's a lot of understanding between the founders." (S5)

From the team members' descriptions, it appeared that they initially felt a sense of loneliness upon entering the accelerator. The contact and networking with other start-ups

in a similar situation helped the teams overcome this subjective feeling, especially at the beginning of the program. The shared focus on overcoming individual challenges in a group created an initial bond between the founders maintained throughout the program. This involved not only supporting each other as a group in overcoming similar challenges, but also psychological and emotional support through a sense of "shared experience" (C2). The founders reported that the purpose of belonging to a group helped them not to lose focus during challenging times, indicating the motivational nature of a cohort, as one of the mentors put it:

> "At some point, there's a curve where it also goes down because then they realize that everything's a little bit more challenging than maybe suspected. And in curves like that, it's always good to meet other like-minded people and realize they're working too and they all have the same problem [...] this provides a very supportive platform for a team." (M)

Overall, the founders' responses indicated that the need for belonging reinforced the ongoing process of relationship building and the intensity of bonds throughout the program.

General overlapping areas

From the team members' descriptions, the degree of *general overlap* emerged as another key reason for reaching out to founders from other teams and similarly informed the intensity of interactions throughout the program. *General overlap* includes common issues start-ups face at the beginning of the founding process, such as how to "approach investors" (S1), what "hiring strategy" (C2) to follow, or how to set up the "fundraising" (C2) process. These types of similarities were cited as a good starting point for contacting other companies working on similar issues or who had previous experience in particular areas. Founders perceived sharing common start-up issues with other teams as helpful in benefitting from each other's experiences, as illustrated by one founder: "You go through certain topics, which every company goes through in the start-up phase, more quickly [...] and often more cost-effectively" (S8). One of the mentors pointed out that early-stage start-ups, in particular, seem to benefit from collaborating with other founders, as the problems "are all still relatively similar" (O-M), in contrast to later-stage start-ups, which require increasingly individualized assistance. In particular, the fact that most teams were in their early stages and thus faced similar challenges increased their willingness to reach out to other teams and talk openly about these issues. In addition, one founder noted that he was initially more attracted to founders who were similar to him in terms of their field of expertise: "Of course, it could be that I unconsciously focus more on these people right from the start" (S8). In general, it turned out that overlapping subject areas or similar areas of expertise significantly influenced the team's willingness to approach each other. Appendix 10, 11, and 12 provide illustrative quotes of the different *network-stimulating factors*.

4.3. Network-activating factors

Network-activating factors form the second category of inherent motivations for founders to form networks with peers. In contrast to *network-stimulating factors*, focusing on intrinsic motivations, network-activating factors describe strategic drivers for participation in network activities beyond the initial contact. These can be divided into: (1) *Anticipated benefit,* (2) *access to external network,* and (3) *knowledge exchange*.

Anticipated benefit

The *anticipated benefit* describes the willingness of the individual founder to build connections with other founders based on the perceived outcome of each relationship. Some founders reported building relationships with their peers without ulterior motives and offering frequent assistance without expecting anything in return, which one of the coaches defined as a "paying it forward" (C2) or "give first" (C2) mentality. This attitude was reflected, for example, in founders forwarding contacts to investors, sharing relevant templates, or providing targeted assistance for business-related topics. One founder shared an example of providing specialized support to another team that was less knowledgeable in a particular area: "We helped clear up the myth of T&Cs simply because we hired costly lawyers who solved the problem for us. And the start-up that asked us how we did the T&Cs now benefits because they can tell their lawyers much more detail about how they want it" (S6). In particular, respondents' statements indicate that start-ups were often willing to share their expertise with others who had less experience in certain areas without necessarily asking for compensation in return.

In contrast, there were cases in which founders only entered into a relationship when the mutual benefit of that relationship was evident, as one of the founders indicated: "They also know that they will benefit if they ever need something that you also help them in return" (S7). Knowing you will get something back indicates a certain level of trust, which seemed to be a fundamental requirement for the development of the relationship, as one of the mentors stated: "In the first phase of relationship building, trust has to be built. Trust as to whether the relationship is somehow reciprocal, that is, whether both benefit from each other, like some kind of win-win" (O-M). This form of reciprocal relationship was evident in internal program sessions in which teams gave presentations to each other and shared content or experiences they found valuable with the rest of the cohort. One of the mentors felt that this kind of connection not only helped founders inspire each other but to tackle certain "questions, topics, problems or challenges as a group" (M). Overall, the interviewees reported several scenarios in which start-ups had benefited from mutually supporting each other. For example, one of the founders of a hardware company mentioned

having helped with the installation of a hardware device for a software-oriented start-up. In return, they received advice on logistical operations, an area within which the other team was comparatively more experienced. It was interesting to learn that the idea of collaboration between teams seemed to be very strong within the cohort, which could be explained by the fact that the founders did not make any statements about competitive behavior throughout the program: "I think once you're an entrepreneur, you're part of a community and this is maybe not the last start-up I do. Maybe I'll do another one. Maybe it's with some of them, maybe not. I see it more as a community, you give and take" (S5). An additional example of the overall cooperative nature of the cohort was found in one of the other founder's narratives: "Some of them had finished pre-seeding right before us. [...] We then benefited from their pre-seeding experience. [...] We have adopted a lot of things that were recommended to us and have also recommended a lot of these things [to the other founders]" (A1). This example suggests that a reciprocal relationship may not necessarily be bilateral, meaning that knowledge was not inherently shared between two parties but at different levels within the cohort.

Access to external network

As an additional reason for networking with other start-ups in the program, founders expressed an interest in expanding their existing portfolio of relevant business contacts. Making new connections and gaining access to a broader network is usually considered one of the essential values of accelerators. Like the accelerator being an intermediary to the regional ecosystem, founders acted as brokers to each other's professional network contacts, such as customers, suppliers, investors, or other business-related contacts. In this regard, founders agreed that reciprocal "referrals" (C2) or "warm introductions" (S4) provided an efficient way to attract potential network contacts and facilitated access to relevant experts:

> "In the end, it's all about connections and about networks in the start-up world. So if you're a founder, the more connections you have, the better off you are. I think it's very cool that now, we know 10 to 12 different start-ups, who I hope will be in business for long, and then you can multiply or expand your network exponentially through others. It's really cool." (S5)

However, as most of the founders were still building on a relatively limited professional network, additional support from the accelerator team was considered relevant in this context to effectively connect the start-ups to the regional ecosystem.

Knowledge exchange

In addition to facilitating network expansion, founders reported the active use of concentrated expertise as one of the key factors in seeking networks with other founders. One of the informants reported initiating a regular roundtable for CTOs to discuss technology-related topics with founders from different teams to benefit from each other's experiences and expertise. Another founder explained that he was particularly looking to connect with teams that were slightly advanced in terms of maturity and expertise to adopt best practices and learn from their experience and insights. Interest in the experience and knowledge of later-stage start-ups included building relationships with alumni start-ups, as one informant reported: "You also can meet the start-ups from previous batches. [...] I think that's really nice because you get that expertise in a way, of people that have been in the same shoes that you are and then you just see them like one or two years later" (S7). In addition to bilateral technical dialogue, founders also cited the importance of networking with other founders for obtaining general information about relevant business tools, key channels for finding investors, or best practices for investment strategies. Similarly, founders reported formulating new use cases or improving their products or services through explicit ideas and tips from other start-ups within the cohort. One founder, for example, recounted a case in which another start-up made him aware of the applicability of his product in a new industry: "I didn't know they had this problem. And that's how we came up with this new solution" (S3).

Overall, the exchange between the teams seemed to be very open and cooperative throughout the program, which could be related to the fact that the teams did not see each other as competitors, allowing for an open exchange of experiences with other teams. A key benefit of sharing ideas at a professional level was learning from each other's mistakes or building on each other's experiences, to quote one of the informants: "After all, they all have similar challenges and everyone has already learned and tried something different and can pass it on. And then the others can learn from it much faster than if everyone has to make the same mistakes again" (MP). Illustrative quotes of *network-activating factors* are summarized in Appendix 13, 14, and 15.

4.4. Network-maintaining factors

Regarding the fact that the program was limited to three months, I was especially interested in the founders' intention to continue using their network relations established throughout the program. Since the second interview phase took place shortly after the end of the program, the data could only be based on assumptions made by the founders. However, I was able to gain further insight into the topic of relationship maintenance through additional interviews with alumni founders and informants from the accelerator team who had already supported several cohorts in the past. Overall, it became apparent that there were significant differences

in the founders' intentions and inherent reasons for maintaining their network, from personal interest to rational or strategic motives.

Network maintenance

The founders' personal interest in *network maintenance* was expressed in particular in their consideration of moving into a shared office with other founding teams after the program. One alumnus mentioned a few "intense relationships with start-ups" (A1) from his cohort, with whom he still regularly exchanges information, provides support, or passes on valuable contacts from his network. One of the mentors, who had worked in the accelerator for several years, reported some very close relationships between individual teams: "Even one or two years after the program, some of them are in daily contact. I know of one example, where five start-ups from one batch moved into an office space afterward because they wanted to continue to have this proximity to the other teams" (O-M). Some of the start-ups reported taking over the planning of networking events beyond the program to keep up with the other teams' journeys. In this context, one founder of the current cohort shared the idea of setting up a regular virtual coffee meeting to keep each other up to date on current topics and challenges, similar to the biweekly update sessions they took part in during the program. Other start-ups, in contrast, showed less incentive to actively maintain relationships but still expressed the added value of established relationships with other teams: "You know, these people are the future. So the fact that we know each other and we can reach out to each other, that's already a lot. We don't have to be best friends. But the fact [...] that we feel free to reach out, I think matters" (S5). Still others felt that without a given structure by the accelerator team, post-program exchanges would most likely become occasional. Finally, some start-ups reasoned that maintaining relationships with other founders was based on rational considerations, such as emerging collaborations, partnerships with start-ups, or perceived cost savings from sharing an office space. Appendix 16 summarizes the illustrative quotes of the *network-maintaining factors*.

4.5. Network dynamics

One of the primary goals of my work was to understand the underlying changes in network formation among early-stage founding teams. By interviewing the start-ups at the beginning and end of the program, several distinctive characteristics emerged about the dynamics of relationship formation, which can be further categorized into (1) *changes in personal and professional proximity* and (2) *changes in perceived networking value*.

Changes in personal and professional proximity

Compared to the beginning of the program, start-ups agreed that the frequency of exchanges between teams increased over the course of the three-months program. In particular, the bonds formed between teams at the beginning seemed to be sustained throughout the program. The first encounter between start-ups, in the form of a two-day kick-off event, was perceived by the founders as essential for building an initial bond between the teams that significantly influenced the further course of the relationship. This was also confirmed by one of the coaches, who observed that as start-ups became more familiar with each other over time, they began to "form typical, inner groups of people that like socializing, the ones that like drinking, the ones that go down smoking, the ones that are just talking about software architecture, the ones that are discussing high-level business" (C1). Besides the kick-off event, which constituted a core driver for relationship building and further network development, I could not find any other pivots in the data that significantly shaped the network, pointing to a relatively gradual process of relationship building. One founder noted that interactions between teams became more efficient over time, as people knew more about the other teams and were thus able to express concerns more specifically or reach out to relevant founders.

Regarding the changes in the intensity of relations, founders expressed different opinions. While only a minority of founders did not experience a significant increase in the intensity of relationships over time, most start-ups reported a positive correlation between the frequency of interactions due to professional or personal overlaps and the intensity of these relationships. Founders reported that relationships intensified the more they learned about the start-ups and their individual challenges. In addition, the data showed that interest in networking with other start-ups increased with the frequency of exchanges. This could be due to the fact that the increased interactions revealed commonalities between the teams, which they could build on over time, as one informant verified: "At the beginning, we didn't have a clue who our people are. [...] Given that we had more interactions throughout these eight weeks you do have the feeling you know people" (S5). Interestingly, one of the alumni founders indicated that the relationships in his cohort became looser and more isolated as the program progressed. This might suggest that the strength of the network equally depends on the individual characteristics of relationships among teams and thus of the cohort as a whole.

This is also in line with what one of the coaches reported in terms of ongoing network activity beyond the official program. According to him, sustained exchange seemed to strongly depend on the composition of the cohort and the underlying intensity of relationships between teams. One of the founders expressed a similar view: "I don't know if I would call them my friends [...] you just support each other and you follow journeys because you're on the same path, like a travel buddy. But, I don't know if necessarily after the journey, we will still be in that close contact" (S7). In addition, one member of the accelerator team reported an example of an "extremely close cohort" that met regularly after the program and proactively maintained intensive contact to maintain technical dialogue. Both statements reinforce

the accelerator team's assumptions about the longevity of relationships as a function of relationship intensity as well as the interpersonal bond built during the program. This was also confirmed by an alumnus reporting on his existing relationships with start-ups with whom he had a very personal relationship and who followed a similar business philosophy to his team. One interesting finding was that despite the continuity of the relationship, the frequency of exchanges decreased significantly after the program. This could be partly related to the fact that start-ups spend less time on networking activities after the program or an indication of the need for long-term networking support from the accelerator. In addition, one alumnus reported the decreasing frequency of exchanges after the program due to the lack of proximity to the start-ups.

The most significant differences in relationship development throughout the program were found between on-site and virtual relationships. Overall, both the frequency and intensity of relationships with virtual teams were uniformly classified as generally weaker by the founders.

Perceived changes in interest in exchange

The *perceived changes in interest in exchange* emerged as the second category when examining the dynamics of the network process. This category describes the increasing recognition of network benefits that emerged among start-ups as the program progressed.

> "It was just when we were still in the pre-seed phase, still very early, I wasn't such a networker, I didn't quite realize how important networks actually are, but then in the stage where we are now, pre-market. That's an optimal place to connect about the same topics that you have in the start-up." (A3)

When asked about the founders' primary motivation for participating in the accelerator program, networking with other founders was rarely mentioned, as one of the founders' statements illustrates: "I didn't mention it [...] because I think it's lower on the priority list. That's not to say it's completely irrelevant, but it certainly hasn't been a deciding factor in applying for [X Combinator]" (S2). Similar statements emerged from the interviews with other founders, who also rated the aspect of the founder network as less relevant. This was partly explained by the lack of knowledge regarding other teams' business models and thematic challenges. In the second round of interviews, the start-ups rated the added value of the network comparatively higher than at the beginning of the program. Founders stated that their interest in the network changed over time when they recognized the added value of exchanging ideas and benefitting from each other's know-how. Informants on the accelerator team also confirmed that the founder's awareness of the value of community was typically not present in the early stages of a start-up. However, the managing partner reported that at the end of the three months, founders frequently named the peer network as the most important added value of the program. A possible explanation for this could be that teams seem to be initially unaware of the thematic overlap with other start-ups and only realize the potential of mutual support over time. Therefore, increasing knowledge about the teams can be classified as a crucial factor influencing the personal perception of the founder network's value. Illustrative quotes of the *network dynamics* can be found in Appendix 17 and 18.

Having explored (1) why and how entrepreneurs interact with other start-ups and (2) how the nature of relationships may change throughout the program, I now elaborate on the role of the accelerator in the context of start-up relationship formation to report on the necessary framework conditions and strategic mechanisms and how these relate to the outcomes of founders' dynamic demands. This serves to answer my third research question: *How does the involvement of accelerators impact the process of relationship building between founders?* As can be seen from my data, accelerators can take on two primary roles. The role of the *accelerator as a boundary system* deals with the underlying conditions necessary for the possibility of network formation, while the role of the *accelerator as a mediator* derives from the practical strategies and reactive activities of the accelerator to the demands of the start-ups.

4.6. Role of the accelerator as a boundary system

Provision of necessary framework conditions

> "We are the orchestrators. So I mean, if we're not there, how are they going to meet, right? We're going to put the excuse event where everybody can join. Who's gonna show them that there are more teams coming, stepping on their steps and standing on their shoulders. What's gonna motivate them to keep saying we were part of [X Combinator]. [...] So we want to make it present because if the incubator is gone [...] who is gonna put them together - nobody." (C1)

The entrepreneurs interviewed agreed that the existence of the accelerator and thus the provision of the framework conditions, including the accelerator environment, was an indispensable prerequisite for forming a network among founders. With regard to the provision of the platform conditions accelerators seek to offer, one of the start-ups' coaches confirmed that the program constituted the main reason for the start-ups to come together. "Without [...] this program, they would have no reason to come together in the first place. So basically, the relationship wouldn't exist at all" (C2). In this context, however, it was also emphasized that the given structure of the accelerator and the involvement in relationship building should be designed in a way that "it is easy for people to interact" (S5) but still allows room for

flexibility and proactivity on the part of the start-ups (S9). This flexibility seems to be provided by the openly designed accelerator environment. In particular, the shared office space offered the start-ups an unconventional atmosphere that enabled the teams to exchange ideas outside the structured program content. When asked to what extent there were differences between the start-ups' on-site relationships, most respondents pointed to the closer relationships and more frequent interactions with their immediate seat neighbors. At the same time, the open environment encouraged teams to occasionally meet with different founders, "join in on a conversation" (S8) or spend time together "beyond their working hours" (C1). Finally, the accelerator setting conveyed an inspiring start-up atmosphere fostering motivation and inspiration among the teams: "It's super important [...] just for motivation or inspiration, it's also important to see what the other founders have achieved" (C1). Another informant confirmed this view: "Well, I think it's, you get this feeling that people come early, stay late. I think you get this feeling that people are working hard, and then you are also motivated to work harder [...] that's a positive thing" (S5). Overall, the open-space climate allowed the teams to follow the progress and daily activities of the other founders, which ultimately increased overall "motivation" (A3), kept "morale" (A3) high and allowed new "perspectives" (S9) to emerge.

Finally, teams indicated that the structure at the beginning of the program was critical to ensure that start-ups had an "initial incentive to connect" (S7), which would "simply take way more time" (A1) to achieve naturally. At the same time, accelerator support still seems to be important later in the program, after founders expressed that they were often overwhelmed with the increasing "start-up hype" related to the maturation of their start-up. However, beyond providing the networking platform and associated structure, one of the start-ups' mentors questioned the program's role in terms of its impact on "how relationships are shaped in depth" (O-M).

Program-specific requirements

An important aspect of examining the role of the accelerator is the program-specific selection process and thus the formation of the cohort to assemble a specific group of founders. The selection process is not only important for obtaining financial sponsorship or increasing overall innovation capacity but also has a significant impact on building collective social capital during and after the program. It is therefore critical in that it can channel early access to new and non-redundant information. Analyzing the data, I found that the level of interaction between teams depends in part on the format of the accelerator, which essentially falls into the following three categories: (1) Niche or generic, (2) stage heterogeneous or homogeneous, and (3) physical or virtual. In addition to these format-related categories, cohort size also appeared to affect the level of team interaction.

The overall difference observed across the first category is reflected in the characteristics of niche accelerators that select teams with a sense of "uniqueness on their way they do things" (C1) and allows to "distinguish them from [...] other companies" (C1) relative to a generic program that does not follow a specific theme. Since the identified case is a technology-focused accelerator, teams are selected based on their technology product or service, combining different industries such as DeepTech, MedTech, Robotics, or AI, as well as different focuses in terms of hardware or software. In this context, one of the managing partners of the program explicitly pointed out the importance of technological heterogeneity, which was deliberately sought to help teams "get out of their bubble, talk to other teams" (MP) and "broaden their horizons" (MP) in terms of expertise and topics. One of the mentors also emphasized the emergence of natural "variance" (M), which, in addition to content-related diversity, can also be achieved through a heterogeneous "personal composition" (M) of teams, bringing together "more experiences, characteristics, ideas, motives, and competencies" (M) and raising the overall learning curve across teams. Surprisingly, unlike what was mentioned among informants of the accelerator team, some start-ups did not perceive this type of selection strategy as conducive to relationship building, mutual interaction, and the emergence of synergy, as put by one of the founders: "For us personally, it would have been better if there had still been companies operating in a similar field to ours. Either in terms of the market or the technology" (S2). He further argued that "if, for example, we had another 14 game studios here, neuro tech or any AI companies that are much closer in terms of technology or content to what we do, then there would certainly be some kind of cooperation that could be pursued further beyond the program" (S2). Another founder agreed with this view due to lower overall "strategic engagement" (S6) among start-ups but noted that having too many similarities between teams would "probably make the exchange less open" (S6). Overall, however, a clear tendency emerged between the different views. The diversity of the teams in terms of market and industry proves to be beneficial in terms of an "open exchange" (S6), reduces the potential of "portfolio conflicts" in the form of competitive thoughts among the teams, and enables the flow of new information. Simultaneously, however, there has to be some form of overlap in terms of similar challenges and issues to be able to discuss general topics on an abstract level and to benefit from the experiences and feedback of multiple "sparring partners" (S9). This goes hand in hand with what founders noted when asked about the impact of different maturity of teams. In particular, the similarity in needs of the start-ups in their early stages allows for an open exchange on industry-independent topics, as one of the founders noted: "For our start-up, I think it was helpful the brainstorming sessions with other start-ups on funding strategy and all of those things. Because then you kind of at the same stage in your start-up and just getting like another perspective [...] and just really chatting openly about it to someone else that is also kind of in your shoes and understands what you're actually going at" (S5). From a strategic point of view, the accelerator team also ensured that all groups were, on average, at a similar stage by delib-

erately hosting "a few earlier-stage teams" (MP) and "a few later-stage teams" (MP) to encourage the flow of knowledge and experience among them.

In addition to the effects of industry and phase-related factors, the data analysis revealed differences in the intensity of relationships and interactions depending on the physical proximity or distance of teams. Overall, sharing a physical space resulted in more frequent and intense interactions between founders on-site than between founders who only participated in the program virtually. In particular, virtual participation prevented unconventional and unstructured opportunities for exchanges in the hallway, at "the coffee machine" (S1), or over "lunch" (S3), that only happened between the on-site teams, through which more informal relationships could be established. The teams that were primarily on-site reported that they had few interactions with the virtual teams because they were simply not "on top of [their] minds" (S5), whereas the connections with the on-site teams became "much closer, more collegial and friendly" (S2) over the course of the program. One founder, who was only able to participate virtually, lamented the lack of closeness to the other teams: "They have these jokes and stuff where you just notice they've spent time together [...] So I think definitely virtual interactions are not as strong for relationship building. [...] I do see it as a disadvantage that I'm not here" (S7). Also, from a psychological perspective, the on-site interactions seemed to significantly impact interpersonal bonding which, according to the founders, had a considerable effect on the strength and maintenance of relationships after the program. In this regard, the accelerator team stressed that a personal bond between participants needed to be established early on to counteract the disadvantages of a virtual setting as efficiently as possible. X Combinator thus introduced an obligatory offsite event at the beginning of the program exclusively to facilitate networking among the start-ups: "Once you've established this bond, which usually works out quite well in two days, they'll stay in touch virtually as well. But you need that in-person contact once, especially at the beginning of the program" (MP). One of the mentors agreed that the development of relationships in a virtual environment would "take much, much longer" (O-M) and make the exchange between teams "less intense" (O-M) than in a physical environment. This highlights the accelerator's function as a broker between start-ups, facilitating networking and exchange, especially between virtual teams.

Finally, the size of the cohort was mentioned as crucial for the exchange, which was summarized by one of the mentors as follows:

> "In my opinion, there is a pareto optimum, which means that too large does not work, because the teams then simply lose the overview of who is doing what, who is facing which challenges, whom can I approach with what [...] and the other side is, if the batch is too small, then you simply have too little overlap of topics, and challenges. In other words, size plays a very important role in my opinion." (O-M)

At the time of the survey, the accelerator comprised 15 start-ups, half of which participated remotely and the other half on-site. When start-ups were asked how satisfied they were with the current size of the cohort, most of them agreed that a larger group could lead to "more professional overlap" (S4), while making it more difficult to "get in touch with everyone" (S2).

Characteristics of the cohort

Consistent with the selection of teams by industry or stage is the *characterization of the cohort*, which can substantially influence the intensity of interaction between founders. This second-order concept essentially comprises the personal attitudes and founder types and the associated proactivity in terms of relationship building with other founders or founding teams. The data suggest that the extent and depth to which connections are formed depends in part on the founders' personalities and interest in other start-ups, as one of the coaches expressed as follows: "You just know that if you hire the right people, the place will run. If you hire the wrong people, it won't run" (C2). Whether a founder is "socially outgoing [or] sharing" (C2) cannot be measured by specific rational criteria, but much more on "human intuition or empathy" (C2) on the part of the accelerator team, which defines the personalities of the founders as a key selection criteria for participation. In this context, it is necessary "just to [have] a few people in the batch who drive [social events]" (MP). Not only the motivation to interact socially with each other was considered a key role, but also the founders' willingness to exchange ideas and share their knowledge with other teams openly. One start-up also noted that in each cohort, "there is always someone in the start-up [who] is responsible for networking and just likes to get everyone connected and likes to address everyone" (A3). The accelerator plays a comparatively subordinate role, as illustrated by the statement of one coach: "So it varies a lot, we just offer the possibility for them to do it and foster these activities. But not everybody's willing to help everybody, but we try to make it happen" (C1). This suggests that the role of the facilitator is required not only within the accelerator team but also within the cohort to ensure bilateral exchange even without active engagement on the part of the accelerator. A different level of proactivity was also found between on-site and virtual teams, with the main factor for lower proactivity primarily attributed to "physical distance" (S9). In addition, some informants among the start-up teams cited insufficient free time as a barrier to personal engagement and building relationships through own efforts. One of the founders noted, "maybe at some stages, you notice that there's other priorities. For example, if it's a very intense stage for a start-up, you will notice that they are less active in these meetings or contribute a bit less, or don't do the networking sessions or whatever" (S5). This difficulty was also acknowledged on

the part of the accelerator team. According to the feedback of the managing partner, it is essential to "convince the teams that [networking] is worth their time" (MP), which often proves challenging, especially for start-ups at a later stage. Also, in this context, the accelerator team emphasized the importance of early bonding between teams, which allows the "relationship, once established" (O-M), to be "leveraged in the future" (O-M) while minimizing the amount of time spent on networking activities. At the same time, a flexible and time-limited program structure also offers teams the opportunity to participate in networking activities according to their available time in order to make relationship building as efficient as possible. Due to time constraints, many interactions happened "door-to-door" (S2) without start-ups explicitly leaving room for networking activities. In contrast, some start-ups proactively and independently organized social events or activities such as bar evenings or thematic get-togethers outside the accelerator setting.

From the overall results, it can be deduced that the difference in proactivity is not only determined by the motivation and interest of the founders but is also a function of different personality types.

Program elements for network facilitation

In addition to the program-specific requirements and the strategic composition of the cohort, X Combinator offered several measures to make it easier for the founders to engage with each other. For example, there was an online communication tool that teams could use to chat with each other or ask specific questions. This was evaluated as a straightforward way for the teams to communicate with each other. In addition to the program-internal channel, there was also a separate mobile application that contained the company profiles of the individual teams, which could be accessed "if [the teams] were looking for something particular" (S8). In addition to listing the company profiles, the accelerator also ensured that teams were regularly informed about the individual challenges and current status of the other founders in bi-weekly stand-up meetings: "So once they are selected, we make them interact more with each other, we have different parts of the program including a presentation from the start-ups to the start-ups, they have the stand-up, they talk to one another" (C1). One of the mentors added that it was essential to introduce the teams to each other: "If the teams know what challenges the individual start-ups are currently facing, they can better assess what specific questions they need to approach the teams with" (O-M). The founders' feedback on the regular update sessions varied. Some of the informants found it helpful in so far as "to better understand the idea of other start-ups" (S10). Others, however, criticized the lack of time in the sessions as well as a pure "working through and reporting things" (S1), which prevented the teams from "dealing with [the problems] in detail" (A2). The lack of structure on the part of the accelerator suggests that these sessions were primarily for the start-ups to help and proactively reach out to each other, which was confirmed by one of the founders: "If you think you can use or need support, you can just reach out to the pool [of founders], do the deep dive, and explicitly ask for [help]" (S6). Overall, start-ups perceived the regular exchange sessions as "good for building relationships" (S7) and allowed teams to share ideas and provide targeted support.

In addition to the stand-ups, workshop formats provided another starting point for founders "to open up, present their challenges to each other, and thus build relationships much more quickly" (O-M). One X Combinator coach reported that such sessions also serve as a foundation for further conversations about the topics discussed: "It's like going to the movies or meeting friends and then talking about it" (C1). Another informant of the accelerator team emphasized the importance of interactive workshops in strengthening relationships. He argued that workshops provided a setting to discuss intimate topics that "probably don't get talked about as much at a pizza night" (M) but are "crucial for further exchange" (M). He explained that superficial topics are usually discussed in the context of unconventional "pizza or beer nights" (M) while addressing deeper issues requires mentor guidance to truly create "value" (M) and "connection" (M) between teams. In particular, sessions with interactive elements such as role-playing provided a way for teams to get closer and "build a personal relationship" (S8). As one of the founders pointed out: "On the first day, one will play a role, the other person would play a different role [...] you get to meet each other in a completely different setting" (S5). This type of interaction encouraged the start-up teams to not only "step out of their comfort zone" (S8) but to solve tasks collaboratively and thus strengthen their personal bond. Illustrative quotes of the *role of the accelerator as a boundary system* are provided in Appendix 19, 20, 21, and 22.

4.7. Role of the accelerator as a mediator

In addition to the necessary but rather passive *role of the accelerator as a boundary system*, a closer look at the data revealed an active role on the part of the accelerator in strategically shaping the network and actively responding to the needs of the start-ups. As evident in my data structure, the *role of the accelerator as a mediator* underlies four second-order themes: (1) *Organization and management of networking events, (2) socialization efforts between start-ups*, (3) *emphasis on network relevance and community* and *(4) strategic design and structure of the network*.

Socialization efforts between start-ups

Socialization between start-ups describes measures to promote personal interactions between the teams, intended to draw the start-ups closer together and break down initial barriers. Ongoing socialization is an integral part of the program and seemed particularly important at the start of the program. Socialization is not only about introducing the start-ups and each other' challenges, but also about spending time together over an extended period of time to "build a community" (O-M) and "create a bond within the group" (O-M). The

teams should be given the opportunity to get to know each other, find out what they have in common, and overcome any fears of contact, as one of the founders described: "What was very good about the program is that the whole event began with this offsite. First of all, there was a very informal atmosphere where everyone could meet on an interpersonal level. And from then on, 80% of the inhibitions were already gone. And people knew each other, they shared the same bedrooms. I think that was a strategically good move by [X Combinator]" (S2) and had a significant impact on how "the dynamic developed over the months" (S2). One of the founders described the event as a type of "speed dating event on a networking level" that helped "break the initial ice" (S2) and "reduce the fear of contact with the other teams tremendously" (S6). The founders agreed on the importance of the "informal, fun setting" (S7) of the kick-off event, which motivated the teams to also talk about "personal things" (S7), to "open up more easily" (S10) and to build "more intense relationships" (A1). One of the founders added that the social meetings at the beginning were important to "better understand the ideas of other start-ups" (S10) in order to reach out to the teams in the further course.

Overall, the data suggest that the progression of relationship intensity and frequency of interactions over the course of the program can be explained in part by the initial socialization of the start-ups, a point confirmed by one of the mentors: "We found that it took much, much longer when it couldn't happen at Corona times. The exchanges were much, much less intense than in those physical batches" (O-M). Another informant emphasized that, overall, physical proximity to the other teams was crucial in this context: "Again, the fact that they took us away to meet and to create a bond from the start, I think that was crucial [...] that's where everything started. We got to know each other, and then we build upon that through the program" (S5).

Organization and management of networking events

The accelerator's measures of team socialization were closely tied to the data that fall into the category of *organization and management of networking events*. In this context, X Combinator places great emphasis on hosting different types of events specifically designed to build and strengthen the internal network and relationships among founders. These events range from casual social gatherings to structured networking sessions to formal business-related events, as one of the mentors explained: "So our networking events are not necessarily just serious events in which they pitch, or they do some formal activities, we always combine them with social events. Dinner, evenings, pizza evenings, drinks, evenings, games. So there is always a socializing factor so that they can get to trust each other" (C1). Opportunities for founders to interact on a social and unconventional level proved particularly helpful in building personal connections and fostering open exchange. With respect to the value of social interactions, one informant noted: "I think it helps the most when we can talk informally with the start-ups in the evenings, because then everyone is a bit more open than when it's immediately clear that it's about a business topic" (S9). Often, the social gatherings were scheduled after more formal meetings or workshops, so start-ups were encouraged to continue discussing the topics in a more informal setting, as one of the founders reported: "We also do a lot of things together in person, also after the networking sessions, just having a beer and pizza and whatever. [...] I think a lot of personal relationships are built up [...] and you just talk on another level with people" (S7). Furthermore, founders reported that also successes of individual teams were frequently celebrated in a joint unconventional setting when, for example, "a start-up closed a financing round" (A1). Celebrating each other's achievements showed the cohorts' strong community spirit and commitment to motivate and support each other.

In addition to the unscheduled social formats, the X Combinator team organizes a series of mandatory networking events specifically focused on "community building and peer exchange" (MP). In particular, an opening event at the beginning of the program, during which the teams spend two full days outside the accelerator environment, serves as a way to get to know each other and make initial contacts with other founders. As already evident from the previous data, the personal bond between the founders, which is often built during these first days, seems critical to building initial trust and further intensifying relationships throughout the program. In addition, founders have the opportunity to exchange ideas and benefit from the experience and expertise of other founders at more formal events. One such event aims for start-ups to "prepare a presentation [of something] they have found valuable" (C1), "organize a thematic roundtable [...] play a game or [...] anything to promote the community idea" (MP). This approach was consistently found to be particularly helpful by start-ups. Here, too, the active role of the accelerator as a facilitator becomes clear, as one of the founders confirmed: "The incentives were set very clearly, you just had to accept them" (S2).

Emphasis on network relevance and community

In addition to organizing and managing events to actively connect the cohort, a key aspect of the accelerator's role seems to be motivating the teams to independently build relationships with each other during the program. In this context, a key concern of the accelerator was to help the start-ups understand the value of networking with each other from the beginning, thus encouraging proactive exchange between teams: "We always strongly justified the community aspect at the beginning. That's also part of our value proposition that we promote and that we then also emphasize very strongly in the intro events [...]" (O-M). In this sense, the main added value of the communication was to articulate the benefits of a community to the start-ups. One of the coaches emphasized that the value of community "is perceived as relatively irrelevant, at least at the beginning when they haven't experienced it yet" (C1). He went on to say: "We actively encourage

them so we keep them informed on what's happening [...] all the available events and we invite them to participate" (C1). The perceived necessity of encouraging teams to engage in mutual networking may be due to the fact that most teams were participating in an accelerator program for the first time and, therefore, had no prior experience in a similar setting. This was also evident in the participants' statements when asked about their initial reason for participating in the program. Teams that had already been part of a funding program noted the added value of building relationships with other start-ups. In contrast, some teams indicated access to the network of investors, partners, and customers as the primary reason for joining the program while assigning less value to connecting with other start-ups. The managing partner also confirmed the different views on network value among the teams: "At the beginning, you still have to force them to actually exchange information. And that is our role [...] so that they realize the value and then continue from there" (MP). In this context, she also noted the difficulty of fostering virtual teams in the same way: "On-site, if you have to, you can get them off their desks and say, 'so, we're going to have a stand-up and you're going to enjoy that eventually'" (MP), which "is not possible with virtual teams" (MP). Equally important seemed to be the insider role of the mentors and coaches, who worked closely with the teams and, therefore, often recognized "content overlaps" (M) or "synergies" (S10) between the founders much earlier. This was also confirmed by one of the cohort's informants: "It's like, 'oh, you know, I just talked to them', and then you have connections. So I think that the mentors and the coaches, they play a vital part in the formation of the networks" (S7). Fundamentally, however, the accelerator team sought to ensure the start-ups' independence and often only encouraged dialogue between the teams without actively connecting them.

Strategic design and structure of the network

Although start-ups were incentivized to network independently throughout the program, data analysis revealed some actions on the part of the accelerator that can be classified as deliberate *strategic design and structure of the network*. Strategic network structures in this context refer to the accelerator's efforts to connect teams precisely according to their technical expertise, thematic overlap, or experience in order to ensure targeted exchange. As one of the mentors mentioned: "The more precise and better matchmaking takes place, the better challenges can be overcome" (O-M). Matchmaking describes the process of bringing together two or more teams that the accelerator expected to maximize synergy and mutual support. One of the coaches provided a tangible example:

> "I happened to learn that they have [X] as a common investor and that they are both active in the tech industry and [...] could maybe learn from each other or even work together, [...] that's such a great match, [...] and then you get the ball rolling. And then all of a sudden, they're in their little world, and of course, they're now working closer together and regularly talking on the phone." (C2)

He further noted that in some cases, the accelerator's intermediary role may even extend beyond the internal network by facilitating relationships with external founders who are operating in the same industry, offer a similar product or service, or are already at an advanced stage and willing to share their experiences with younger founders. He argues that "proactively identifying the needs of founders and then connecting them with the community in a way that adds value to one or both sides" (C2) is a crucial part of the process. Regarding the efficiency of matchmaking, the founders' feedback was mixed. One founder, in particular, criticized the potential creation of "dependencies" (S4), especially when creating partnerships between start-ups, which he experienced within another accelerator program. In contrast, specific formats that brought together groups of founders with thematic overlaps were felt to be helpful in "facilitating conversations" (S10) and "enabling knowledge transfer" (MP). However, one of the founders felt that thematic exchanges needed to be structured in a way that allowed for "discovering the specific problems as efficiently as possible, defining them and articulating them in an appropriate framework, in order to discuss" (S1) them efficiently.

Matchmaking occurs not only within the current cohort but also across cohorts, allowing younger start-ups to connect with start-ups von previous cohorts. One way to create this exchange is for later founders to share their experiences with current teams through "workshops, talks, or one-on-one mentoring" (C2). One of the founders observed that it's not so much the idea of connecting start-ups at the same stage that adds value, but rather the contact with experienced start-ups "that are maybe a year or two ahead of us, because they went through the things that we are going right now" (S5). He further noted that alumni support would represent a form of "paying back" (S5) toward the accelerator "to help the next generation of founders" (S5), reflecting his intention to stay in touch with the accelerator and the teams even after the program. In this context, however, he stressed the need for efficient design of exchanges, especially in view of the lack of time:

> "In reality, I have a business to run day to day. So this is the top priority, but given that I was helped, I'm more than willing to pay back and as long as it's made easy for us to do so. I think that's maybe what [X Combinator] should focus on. [...] Don't ask me for 10 times one-hour meetings, but actually, invite me at an event and I will then mingle with these people and share everything I know." (S5)

The data showed that founders disagreed on what constituted accessibility to alumni teams. Some founders pointed to the relatively loose contact with alumni and blamed it on

the lack of support from X Combinator, while other founders felt no difficulty making contact on their own. In terms of wanting to get alumni teams more involved in the program, one of the founders had the idea of creating a "space [for alumni]" (S2) to facilitate access to the teams while "serving as motivation" (S2) for the younger start-ups. Overall, most start-ups felt that alumni teams were generally willing to share their experiences with younger founders and were responsive to their questions.

The accelerator team not only provided start-ups with the opportunity for cross-cohort exchange but also emphasized relationship maintenance and follow-up support, as one of the coaches noted: "it poses an opportunity for them to reconnect again. [...] At least we give them an excuse to come by and say hi to each other. [...] So they get to interact for as long as the accelerator is still alive" (C1). Some start-ups frequently remain close to the accelerator after participating in the program by joining another related program, allowing them to maintain contacts more easily. Other start-ups often decide to move into a joint office or to cooperate in some way after finishing the program, although this is mostly arranged bi-laterally between the teams, without active support from the accelerator side. However, alumni find the events organized by the accelerator team helpful, as they provide them with a framework for long-term exchange. After the three-month program, start-ups also have the option of extending their stay at the accelerator and using its facilities for an additional three months. However, the intensity of relationships did not seem to increase significantly in this context, which could be partly explained by the founders' increasingly busy schedules. Appendix 23, 24, 25, and 26 summarizes the illustrative quotes of the *role of the accelerator as a mediator*.

4.8. Propositions and process model development

The results underlie the seven aggregate dimensions presented in the data structure. In essence, they provide insight into (1) how start-ups proactively develop networks with other founders and the motivations underlying network development, (2) the ways in which different network connections are strategically leveraged and change over time, and (3) the role of the accelerator in network formation to effectively and timely respond to the relevant needs of start-ups. Based on the analyzed results, I derived four main propositions.

Dynamic networking behavior

As I sought to understand with whom, when, and for what purpose the entrepreneurial teams networked during the interviews, it became clear that the configuration of relationships and interactions over the course of the program was related to each team's inherent motives at a particular point in time. Through closer analysis of the interviewees' statements, I found that the process of networking could be broken down into three major phases. As already evident in the data structure, these can be delineated based on *network-stimulating*, *network-activating*, and *network-maintaining* motives, which are determined by the inherently different requirements on the part of the founders. Based on these findings, I argue that

> **Proposition 1** - P1 (Network Stimulation, Network Optimization, Network Maintenance): *Networks among early-stage founders transition through various stages of formation to accommodate shifting organizational demands and challenges over time.*

Strategic Network Management

Furthermore, the results of the data analysis suggest that the accelerator plays the role of a powerful connector between the start-ups in an ecosystem by creating a protected environment that encourages actors to collaborate, share knowledge, and exchange information. What looks like a passive assembly of individual actors is based on the idea of active cognitive navigation between complementary know-how and behaviors that holistically strengthens the overall value of entrepreneurial connectivity.

The data suggest that the upstream selection process, and thus the strategic formation of a cohort of founders, enables the efficient coordination of complementary knowledge and skills. In addition, the accelerator's case-by-case assessment of whether a team fits into the overall founder network strengthens trust among founders and a sense of collective identity beyond the accelerator's boundaries. From these results, I deduce that

> **Proposition** - P2a (Network Composition): *Strategic network composition is an effective tool to filter qualified start-ups based on their network fit to ensure targeted interaction among founders and catalyze long-term connectivity.*

As described by informants, founders seek to dynamically align their peer network connections with their perceived needs. However, early-stage founders, typically face the difficulty of anticipating strategically valuable relationships based on unpredictable challenges (Engel et al., 2017). At this point, the responsibility shifts to the accelerator, whose insider role can align the interests of one founder with the experiences or capabilities of another. In other words, the accelerator's expertise in strategically connecting start-ups contributes to optimized network formation and thus effective exchange and collaboration among founders. I, therefore, suggest that

> **Proposition** - P2b (Network Orchestration): *Matchmaking between start-ups constitutes a cognitive mechanism for enhancing the efficiency of entrepreneurial networks and ensuring a facilitated flow of information between founders.*

Socialization Incentive

While strategically motivated connections among start-ups are associated with reciprocity in exchange processes, intrinsically motivated connections form the basis for personal ties that create trust and emotional bonding. As the informants' statements indicate, professional-level connections, that is, calculative relationships, emerged temporally only after the start-ups had connected on a personal level during the initial encounter. This suggests that strategically motivated exchange dynamics between start-up members are usually a downstream event in the development process that requires a previously established form of interpersonal bond between the start-ups. Moreover, the data showed that the regularity and intensity of the relationship were based not only on professional compatibility but also on personal connections between the founders. This observation underscores the importance of fostering early social interactions to strengthen interpersonal bonding. These findings lead me to conclude that

> **Proposition** - P3 (Network Activation): *Mediating early network formation among early-stage founders increases the likelihood of interpersonal bonding and thus the frequency and intensity of mutual collaboration over time.*

The propositions derived from the results can be presented as a dynamic input-output model of inter-firm network formation. The process model shown in Figure 4 depicts the various phases of network formation in chronological order. In practice, the boundaries of the phases may not be as discrete, suggesting possible shifts or adjustments depending on the particular context. For example, relationships at the business level are not necessarily limited to the network maintenance phase but may already be formed earlier in the process.

The interaction dynamics between the founding teams at the different networking levels form the core of the model. *Start-up inputs* describe the intrinsically and extrinsically motivated incentives of start-ups (*initiators*) to build connections with other founders. *Accelerator inputs* represent the strategic mechanisms (*initiatives*) designed in response to the start-ups' expected needs and specific circumstances. Finally, the *accelerator output* comprises the corresponding outcome obtained through the different types of connections within the overall founder network.

The applicability of the model can be illustrated with a simple example (highlighted in blue in the model): At the beginning of the program (*stimulation phase*), most start-ups are confronted with a new situation in which they primarily seek contact with like-minded people in their immediate environment (*start-up input*). Accordingly, strategically initiated socialization, in the underlying case, the opening event at the beginning of the program (*accelerator input*), helps to create early interpersonal connections between the founders. Ultimately, these initial connections create an emotional bond and mutual trust between teams (*accelerator output*) that builds the foundation for deeper connectivity.

The theoretical and practical implications of the theses developed, and the process model will be examined in more detail below.

5. Discussion

The goal of this thesis was to develop a deeper understanding of different types of relationships between early-stage founders and the associated *dynamic changes over time: How do early-stage start-ups form, use, and develop relationships with other founders? How does the importance of different relationships change in light of the dynamic nature of entrepreneurial activity? And how does the involvement of accelerators impact the process of relationship building between founders?* This research not only examines the procedural dynamics in developing relationships between start-ups but simultaneously challenges the effectiveness of contemporary support programs for entrepreneurs.

Positioning the accelerator at the meso-level of analysis allowed me to examine its intermediary role between the network's actors. Studying a realistic case in its natural setting was consistent with the overall goal of studying the process and associated temporal dynamics of inter-organizational relationship formation among start-ups over time (Denzin & Lincoln, 2011). Semi-structured interviews with founders and representatives of the accelerator team formed the basis for an inductive, qualitative research study.

The developed theoretical process model represents a significant contribution to the vast literature on entrepreneurial networks and provides valuable guidance to the new generation of ecosystem designers, policymakers, and prospective entrepreneurs. In the following section, I outline the implications and assumptions derived from the results before presenting limitations and suggestions for further research.

5.1. Theoretical implications

The results of the underlying study provide a number of insightful findings for the broader academic literature on entrepreneurial networks and the nascent field of accelerator research. Overall, they expand the definition and understanding of social founder relationships and illustrates the complementarity of networks in entrepreneurial processes.

First, they conceptualize entrepreneurial networking as an inherently adaptive behavior of individual actors along the entrepreneurial process. Much of the prevailing empirical research emphasizes the centrality of networks in all stages of the entrepreneurial process (Stuart & Sorenson, 2007). However, most of the literature on entrepreneurial networks has focused on defining the structural characteristics of entrepreneurial networks as an outcome of entrepreneurial activity rather than considering entrepreneurial networking as the activity itself (Engel et al., 2017). As previously mentioned, academic research has been primarily devoted to the

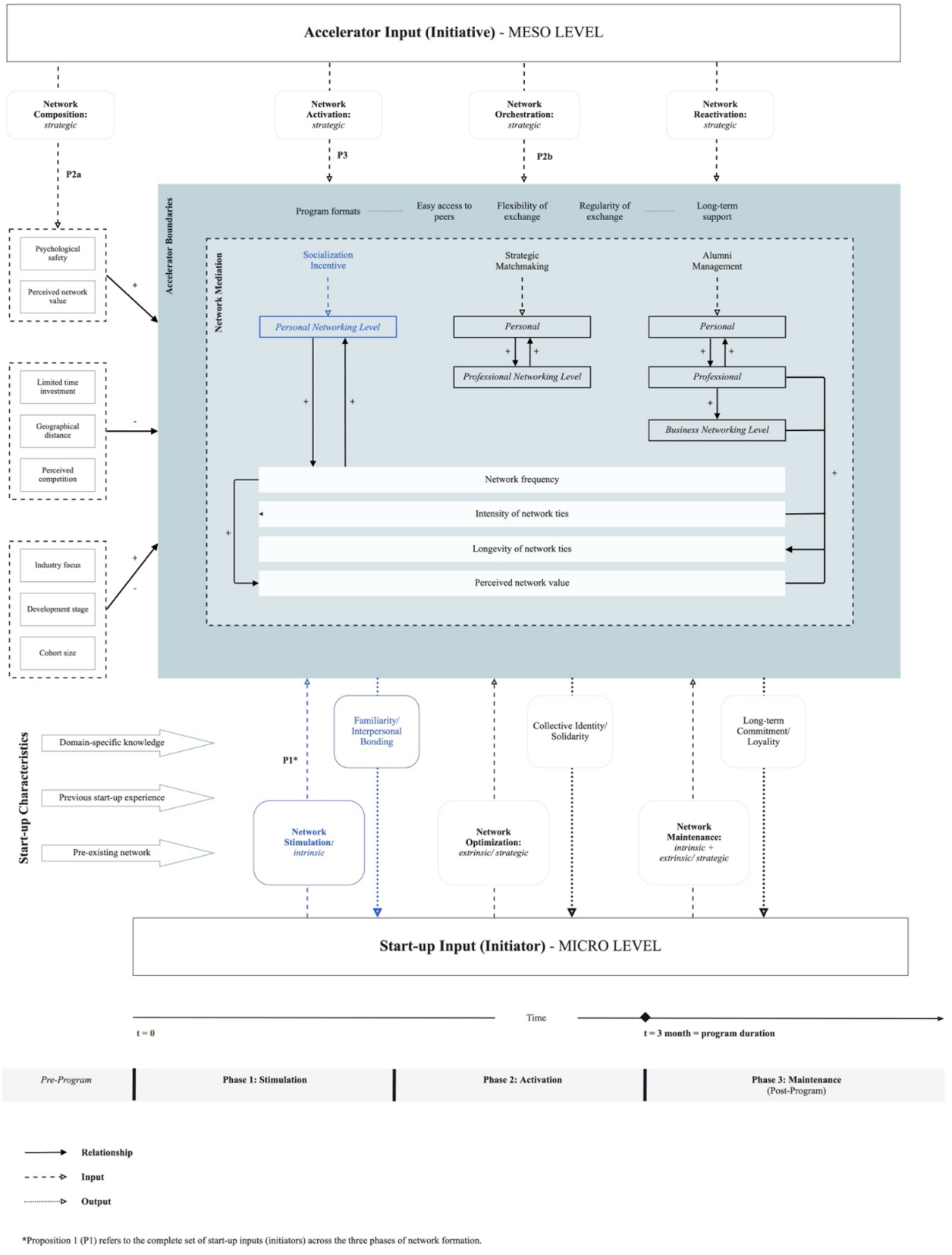


Figure 4: Process model - Dynamic input/output model of inter-organizational network formation (Source: Own illustration)

what, that is, the structural characteristics of entrepreneurial networks, while relatively little comprehensive knowledge exists of the *why*, that is, procedural elements and behaviors underlying entrepreneurial network activity (Evald et al., 2006).

Especially in the entrepreneurial context, however, it seems important to not only understand how nascent companies gain initial access to established networks but also how they proactively build effective networks from scratch. Therefore, in contrast to previous research, this study refrains from viewing network structures as exogenously given but rather characterizes entrepreneurs as active actors within a social network proactively seeking connectivity with their peers. Consequently, the active role of the entrepreneur in connecting with other founders takes a central role in my study. From the perspective of the active entrepreneur, the proposed theoretical framework should, therefore, be able to answer key questions about strategic goals, required resources, and potential uncertainties, but also provide insights on how to leverage and further develop the network. The model not only provides a basic representation of the various input and output factors guiding the different social interaction mechanisms but also extends our theoretical understanding of potential variability within founders' interaction mechanisms as well as along the stages of network formation.

Second, the findings illuminate and explore the flexible forms of inter-firm relationships that emerge in response to changing requirements at different stages of the development process. Moving away from a structural view and adopting a procedural perspective was necessary not only to understand the inherent reasons and motives of emerging founders for forming networks with their peers but also to gain a more comprehensive picture of the associated dynamic change processes over time. While most theoretical concepts examine networks at a single, fixed point in time (Evald et al., 2006), the underlying study provides new theoretical insights into the dynamic nature of the entrepreneurial network process. The results show that relationships between start-ups evolve organically and change over time in response to the agile nature of entrepreneurial processes and the underlying transformational needs of founders.

In particular, the results contribute to previous theories on entrepreneurial networking by providing a conceptual framework for the different types of relationships between founders in the initial phase of their development, as well as an explanation for the associated change processes in the related structures. In this regard, the study's results reveal the different stages of relationship formation from the first encounter to the end of the program. In addition, they show how the different types of interactions arise throughout the process as well as the associated relationship levels along which founders typically navigate during the course of the program. Based on the study results, the process model not only provides information on the chronological sequence of the networking phases and the different characteristics but also depicts the corresponding content, motives, and outcomes underlying the respective relationships between the actors. Essentially, the results suggest two main building blocks of inter-organizational relationship building: Personal relationships (personal network level) and company-specific relationships (professional and business level). Overall, these findings expand our holistic understanding of the different types of social interaction mechanisms and networking processes within the accelerator environment.

Third, the findings underscore the practical and theoretical importance of the accelerator's role in strategically supporting relationship building among founding teams to ensure optimized exchange and collaboration. The approach of describing the accelerator as a closed social network allowed the phenomenon to be studied in terms of underlying dynamics rather than specific structural components and design elements. Certainly, one of the greatest difficulties in forming social networks is the ability of nascent entrepreneurs to identify and enter into appropriately relevant relationships for the optimized development of their business. Compared to the ubiquitous role of accelerators in connecting young founders with relevant contacts in the regional ecosystem, the role of the accelerator also includes its inherent ability to catalyze internal network connections between teams of founders. While recent research on accelerators has provided evidence of the potential added value of collaborative peer environments (Cohen, Fehder, et al., 2019), no clear evidence exists on whether these benefits were achieved through proactive collaborative efforts by founders or strategic management of reciprocal exchanges and interactions on the part of the accelerator. Hence, this study sought to resolve this particular ambiguity by examining the evolution of the founders' relationships over time. Exploring the dynamics of peer interaction within social events allowed me to determine the respective conditions under which the accelerator's role became evident.

Findings show, how accelerators can serve as platforms for the emergence of start-up networks while also taking an active role in the strategic formation and development of relationships among founding teams. Drawing on this insight, the process model illustrates this duality of interaction dynamics by mapping the role of the founder as the initiator (pursuing a specific goal, interest, or need) and the role of the accelerator as a reactive initiative (responding to the founder's goal, interest, or need). In other words, it becomes clear that the willingness and incentive for mutual interaction must exist on the part of the founding teams and cannot be imposed by the accelerator. In contrast, however, it is the responsibility of the accelerator to identify or anticipate the specific needs of the founder to provide appropriate support initiatives or encourage mutual exchange.

Finally, this theoretical account is closely related to academic research on entrepreneurs' motives, incentives, and behavioral strategies in building social networks. Previous theories of entrepreneurship and network research differ, among other things, in their interpretation of entrepreneurial action. While some studies explain network configuration in terms of rational agency (Miller, 2007), others view rela-

tionship formation in the context of entrepreneurial uncertainty (Alvarez & Barney, 2007; Engel et al., 2017). Rather than committing to one of these views, this study combines both rationality and uncertainty in entrepreneurial behavior. Looking at the results and the derived process model, this approach becomes clear: Peer networks are formed partly out of strategic motives, but cannot be implemented completely rationally due to the uncertainty inherent in the entrepreneurial process (Engel et al., 2017). Again, this highlights the role of the accelerator as an insider helping to bridge or mitigate entrepreneurial uncertainty through strategic matchmaking between founders. This is not only a novel but an important theoretical insight within accelerator research as well as a theoretical answer to the question of how to deal with entrepreneurial uncertainty in the context of social networking.

Taken together, on the one hand, the underlying study findings and the resulting process model highlight the importance of examining motivational aspects of relationship formation on the part of the entrepreneurial actors to appropriately design the content of accelerator programs and strategically manage relationships between founders. On the other hand, the different forms of social interaction mechanisms over time inform our holistic understanding of the dynamic processes within founder networks and encourage additional research on entrepreneurs' social interaction mechanisms.

5.2. Practical implications

The theoretically developed framework offers valuable practical insights into the dynamic nature and unique characteristics of start-up relationship building. It provides suggestions for structuring and managing accelerators for effectively orchestrating and shaping start-up networks in the future. The findings of this study highlight the importance of social networks among early-stage founders and conceptualize networking as an integral and ongoing entrepreneurial activity throughout the entrepreneurial journey. Building networks with other founders should therefore be considered an inherently adaptive process in order to respond to shifting demands related to the dynamics of entrepreneurial processes, especially with regard to the maturity of the start-up.

In this respect, accelerators should be aware of their critical role as strategic catalysts for the formation of networks between start-ups and the various channels and tools they can use to foster and actively manage collaboration. The overall results show how network configuration becomes more specific as start-ups mature, likely due to the increasingly unique and specialized needs of founders. Therefore, ensuring an ongoing network strategy and design is critical for a successful entrepreneurial career. From a meso-perspective, which means, from the accelerator's point of view, this requires management to redesign or adapt the program to optimize internal connectivity and collaboration among start-ups. It is reasonable to assume that current efforts to promote entrepreneurial progress may not be realizing their full potential in this regard. Therefore, accelerators are well-advised not only to consider the composition of a cohort but also to implement measures and promote activities that support cohesion and the development of social capital among teams. In this regard, program structures and event formats should be designed in a way that allows teams to socialize beyond the regular content sessions. In doing so, the accelerator should integrate the idea of peer-to-peer networking as a core value in the program to promote the longevity of relationships and collaboration between start-ups beyond the program boundaries.

For founders, this study presents start-up networks not only in terms of their role in providing personal and emotional support but also as an important strategy for efficient development. Early-stage start-ups in particular should recognize the value of collaborative networks with other founders to motivate each other, provide professional advice, and gain access to external networks. In this regard, the results underscore the essential role of founders to willingly and actively seek out contact with peers - either directly by reaching out to appropriate parties or indirectly by openly articulating their needs. In practice, this implies that entrepreneurs should engage in ongoing initiatives that facilitate contact with founders and foster long-term connections. In addition, they should maintain a positive attitude toward sharing knowledge and experiences with other founders and adopt an overall cooperative mentality toward the network. However, the role of the committed and motivated founder makes the accelerator a vulnerable and sensitive model as soon as initiative and proactivity on the part of the entrepreneurs are missing. With this knowledge, accelerators are encouraged to integrate collaborative initiatives and continuously incentivize founders to proactively network with their direct environment.

Finally, creating the necessary framework conditions for start-ups is a fundamental part of funding policy. The results provide practical guidelines for policymakers, who are encouraged to evaluate accelerators not only based on their structural elements, but also on their ability to create efficient and long-term founder networks. Specifically, this means that criteria for assessing the efficiency of accelerators should be based on their ability to select relevant participants, develop efficient internal relationships, and coordinate them strategically to build an adequate network of founders. Such evaluation standards accordingly inform both entrepreneurs and the broader entrepreneurial landscape.

5.3. Limitations and future research

Although the underlying findings not only add to the basic understanding of what entrepreneurship is and how it emerges, my study reveals some limitations to consider. Moreover, the fragmented nature and novelty within the topic of founder networks provide insights for further research.

The primary question to be answered in the context of qualitative inductive studies is the extent to which the underlying findings allow for generalizability. According to Gioia et al. (2013), when dealing with a single case study, the issue

is less about generalizability and more about the transferability of the underlying phenomenon, that is, the applicability of the results to other contexts. Based on this assumption, generalizability is not to be understood in the sense of statistical representativeness but rather in the sense of theoretical plausibility. Hence, based on the underlying case study, the question arises to what extent the results are related to larger social contexts so that descriptions and explanations of the smaller social unit can be transferred similarly to a broader domain (Brüsemeister, 2008). Since the chosen accelerator setting represents a unique context, the corresponding results cannot be easily transferred outside this specific setting. Start-ups operating in an accelerator environment are part of a closed and protected world explicitly created to bring them together. Building relationships outside such an artificially created system might be subject to fundamentally different conditions and may require a whole new perspective. Therefore, it seems reasonable to compare the theoretical insights gained on network formation and its significance for founders in the context of the particular ecosystem in which they operate. This should also involve a closer examination of the role of the accelerator in relationship building with that of another type of support institution in order to uncover possible similarities or differences. Although studying the phenomenon allowed to build sufficiently robust theoretical propositions, suggesting transferability to other context (Gioia et al., 2013), social realities are often too complex to be investigated by a single research method (Edmondson & McManus, 2007). Therefore, validating the results against quantitative methods is strongly recommended to assess the quality of my findings and statistically test the proposed associations beyond the scope of this study.

If, however, we considers the accelerator not as an isolated unit but as a self-contained social structure within an overall system, it can be assumed that the internal relationships between the founders and the emerging founder network similarly reflect a part of the overall network. In other words, it is reasonable that the results relating to the formation of relationships between start-ups within an accelerator are nonetheless a result of the overall start-up ecosystem. This assumption suggests that relationships between early-stage start-ups might similarly develop beyond the boundaries of an accelerator. Thus, if networks between start-ups make a critical difference in the formation of a company, it seems essential to understand whether the added value is determined by specific characteristics of the network composition and thus independent of the setting or by the particular environment in which the network is formed. Therefore, future research will need do assess the model's applicability in other settings, particularly outside the accelerator landscape (e.g., across other emerging companies not supported by an accelerator) and beyond early-stage start-ups (e.g., across founding teams of different maturity).

Contextual considerations are closely related to another limitation often encountered in qualitative studies. Generally, researchers often assume that the insights gained from qualitative studies account for the nature of the underlying phenomenon as well. It should be noted, however, that entrepreneurship as a whole produces a wide range of different phenomena that may vary depending on context and individual circumstances (Gartner & Birley, 2002). First of all, entrepreneurship is considered a very heterogeneous phenomenon, which is why actors frequently deal with the same circumstances in different ways (Welter et al., 2017). Thus, we cannot simply assume that entrepreneurs go through the different phases of network formation in the same way (Greve & Salaff, 2003). This is consistent with the fact that the lifecycle of start-ups sometimes exhibit fundamental differences across domains and industries. For example, network requirements certainly differ between nonprofit social enterprises and high-growth start-ups in the technology sector. The phenomenon studied in this research includes processes and characteristics that are common to technology-based start-ups but cannot be simply applied to start-ups in other industries. In order to be able to make accurate statements about the general validity of the underlying results, further research is needed to validate the theoretical results across industries. Beyond that, however, it may be interesting to conduct more in-depth studies within a single industry. If we assume that network relationships offer corresponding advantages in entrepreneurial contexts, the question arises whether all actors in a network find equal conditions and can draw the same output from the network respectively, or whether it is specific characteristics of the individual actor that lead to network-based advantages. Future studies might, therefore, discuss the presented theoretical framework in the context of different institutional, economic and cultural settings to make potential relationships between context, network, and individual network actors more tangible.

Taken together, due to the limited duration founders spend in an accelerator as well as its artificially created environment, it would be naïve to conclude from the particular, that is, the specific case, to the general, that is, the environment of the case and thus to sociocultural rules (Bürsemann, 2008). Certainly, however, accelerators provide an essential cornerstone for initiating exchange and establishing social bonds among founders in the early stages that might equally be leveraged in other contexts. Therefore, the extent that the underlying findings represent the boundaries of best practices for a single accelerator, they should serve to provide both guidance and lessons that can be adopted more broadly. As such, the model developed not only represents a theoretical construct with demonstrable transferability for the subject area studied, but also serves as a foundation for scholars, practitioners, and entrepreneurial stakeholders alike.

6. Conclusion

Developing, promoting, and managing networks is a core element of entrepreneurship. Building efficient relationships and a solid network becomes a key success factor in today's competitive global market. Yet, the potential of collaborative

networks between emerging companies is often underestimated. In this study, I employed a qualitative, inductive approach allowing me to examine processes and mechanisms at multiple levels over time. Specifically, I developed a dynamic process model based on inter-organizational relationships of early-stage start-ups operating in an accelerator environment. Within the socially situated framework of the accelerator, I was able to examine the process elements and interaction dynamics involved in the formation of different types of relationships between nascent founders. The process model developed focuses on the key characteristics and conditions related to inter-firm relationship building. It captures how founders adaptively shape their relationships with peers in light of changing personal and organizational needs while illustrating the intermediary role of the accelerator in strategically connecting founders to foster an effective network and create a cohesive environment. Findings suggests that relationship formation among start-ups is a likely replicable type of process, the underlying dynamic nature of which may be relevant to both nascent founders as well as entrepreneurial support institutions. Novel insights should accordingly be used to guide the design and strategic management of start-up relationships. In order to achieve associated goals, accelerators are required to revise their business model incorporating collaborative initiatives and ongoing network support. Overall, I hope that the underlying findings and theoretical propositions will further stimulate research on the dynamic and interactive nature of inter-organizational networks and encourage dialogue across relevant domains.

References

Acs, Z. J., Autio, E., & Szerb, L. (2014). National systems of entrepreneurship: Measurement issues and policy implications. *Research Policy*, *43*(3), 476–494.

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

Adner, R., & Kapoor, R. (2010). Value creation in innovation ecosystems: How the structure of technological interdependence affects firm performance in new technology generations. *Strategic Management Journal*, *31*(3), 306–333.

Aguinis, H., Gottfredson, R. K., & Joo, H. S. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270–301.

Ahmad, A. J., & Ingle, S. (2011). Relationships matter: Case study of a university campus incubator. *International Journal of Entrepreneurial Behaviour & Research*, *17*(6), 626–644.

Aldrich, H., & Zimmer, C. (1986). Entrepreneurship through social networks. In *The art and science of entrepreneurship* (pp. 3–23). Ballinger Publishing Company.

Alvarez, S. A., & Barney, J. B. (2005). How do entrepreneurs organize firms under conditions of uncertainty? *Journal of Management*, *31*(5), 776–793.

Alvarez, S. A., & Barney, J. B. (2007). Discovery and creation: Alternative theories of entrepreneurial action. *Strategic Entrepreneurship Journal*, *1*(1-2), 11–26.

Baron, R. A., & Markman, G. D. (2003). Beyond social capital: The role of entrepreneurs' social competence in their financial success. *Journal of Business Venturing*, *18*(1), 41–60.

Baum, J. A., Calabrese, T., & Silverman, B. S. (2000). Don't go it alone: Alliance network composition and startups' performance in Canadian biotechnology. *Strategic Management Journal*, *21*(3), 267–294.

Bernard, H. R. (2017). *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman & Littlefield.

Bøllingtoft, A. (2012). The bottom-up business incubator: Leverage to networking and cooperation practices in a self-generated, entrepreneurial-enabled environment. *Technovation*, *32*(5), 304–315.

Bøllingtoft, A., & Ulhøi, J. P. (2005). The networked business incubator: Leveraging entrepreneurial agency? *Journal of Business Venturing*, *20*(2), 265–290.

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management*, *29*(6), 991–1013.

Borgatti, S. P., & Halgin, D. S. (2011). On network theory. *Organization Science*, *22*(5), 1168–1181.

Bourdieu, P. (1983). Ökonomisches Kapital, kulturelles Kapital, soziales Kapital: Soziale Welt. In R. Kreckel (Ed.), *Special Issue on Social Inequalities* (pp. 183–198). Schwartz.

Bourdieu, P. (1985). The forms of capital. In J. G. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). Greenwood.

Brüsemeister, T. (2008). *Qualitative Forschung: Ein Überblick*. VS Verlag für Sozialwissenschaften.

Burns, B. L., Barney, J. B., Angus, R. W., & Herrick, H. N. (2016). Enrolling stakeholders under conditions of risk and uncertainty. *Strategic Entrepreneurship Journal*, *10*(1), 97–106.

Burt, R. S. (1992). *Structural Holes*. Harvard University Press.

Burt, R. S. (1997). The Contingent Value of Social Capital. *Administrative Science Quarterly*, *42*, 339–365.

Burt, R. S. (2001). Structural holes versus network closure as social capital. In N. Lin, K. Cook, & R. S. Burt (Eds.), *Social Capital: Theory and Research* (pp. 31–56). Aldine de Gruyter.

Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, *110*(2), 349–399.

Cassar, G. (2004). The financing of business start-ups. *Journal of Business Venturing*, *19*(2), 261–283.

Cohen, S. L., Bingham, C. B., & Hallen, B. L. (2019). The role of accelerator designs in mitigating bounded rationality in new ventures. *Administrative Science Quarterly*, *64*(4), 810–854.

Cohen, S. L., Fehder, D. C., Hochberg, Y. V., & Murray, F. (2019). The design of startup accelerators. *Research Policy*, *48*(7), 1781–1797.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, *94*, 95–120.

Coleman, J. S. (1990). *Foundations of social theory*. Harvard University Press.

Crawford, G. C., Aguinis, H., Lichtenstein, B., Davidsson, P., & McKelvey, B. (2015). Power law distributions in entrepreneurship: Implications for theory and research. *Journal of Business Venturing*, *30*(5), 696–713.

Davidsson, P. (2016). *Researching entrepreneurship: Conceptualization and design*. Springer.

Davidsson, P., & Honig, B. (2003). The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing*, *18*(3), 301–331.

Davidsson, P., & Wiklund, J. (2001). Levels of analysis in entrepreneurship research: Current research practice and suggestions for the future. *Entrepreneurship Theory & Practice*, *25*, 81–100.

Davis, J. A. (1963). Structural balance, mechanical solidarity, and interpersonal relations. *American Journal of Sociology*, *68*, 444–462.

Denzin, N. K., & Lincoln, Y. S. (2011). *The Sage handbook of qualitative research*. Sage Publications.

Dimov, D. (2011). Grappling with the unbearable elusiveness of entrepreneurial opportunities. *Entrepreneurship Theory and Practice*, *35*(1), 57–81.

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

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

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32.

Engel, Y., Kaandorp, M., & Elfring, T. (2017). Toward a dynamic process model of entrepreneurial networking under uncertainty. *Journal of Business Venturing*, *32*(1), 35–51.

Evald, M. R., Klyver, K. I. M., & Svendsen, S. G. (2006). The changing importance of the strength of ties throughout the entrepreneurial process. *Journal of Enterprising Culture*, *14*(1), 1–26.

Evans, P. (1996). Government action, social capital and development: Reviewing the evidence on synergy. *World Development*, *24*(6), 1119–1132.

Fuchs, S. (2020). *Geltungsbereiche des sozialen Kapitals in Deutschland*. Springer.

Gartner, W. B., & Birley, S. (2002). Introduction to the special issue on qualitative methods in entrepreneurship research. *Journal of Business Venturing*, *17*(5), 387–395.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31.

Glaser, B. G., & Strauss, A. (1967). *The discovery of Grounded Theory*. Aldine Publishing Company.

Glaser, B. G., & Strauss, A. (2008). *The discovery of Grounded Theory: Strategies for qualitative research* (3rd). Aldine Publishing Company.

Goswami, K., Mitchell, J. R., & Bhagavatula, S. (2018). Accelerator expertise: Understanding the intermediary role of accelerators in the development of the Bangalore entrepreneurial ecosystem. *Strategic Entrepreneurship Journal*, *12*(1), 117–150.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Greve, A., & Salaff, J. W. (2003). Social networks and entrepreneurship. *Entrepreneurship Theory and Practice*, *28*(1), 1–22.

Gulati, R. (1998). Alliances and networks. *Strategic Management Journal*, *19*(4), 293–317.

Hallen, B. L., Cohen, S. L., & Bingham, C. B. (2020). Do accelerators work? If so, how? *Organization Science*, *31*(2), 378–414.

Hallen, B. L., & Eisenhardt, K. M. (2012). Catalyzing strategies and efficient tie formation: How entrepreneurial firms obtain investment ties. *Academy of Management Journal*, *55*(1), 35–70.

Hansen, E. L. (1995). Entrepreneurial networks and new organization growth. *Entrepreneurship Theory and Practice*, *19*(4), 7–19.

Hansen, M. T., Chesbrough, H. W., Nohria, N., & Sull, D. N. (2000). Networked incubators. *Harvard Business Review*, *78*(5), 74–84.

Heider, F. (1958). *The Psychology of Interpersonal Relations*. John Wiley & Sons.

Hoang, H., & Antoncic, B. (2003). Network-based research in entrepreneurship: A critical review. *Journal of Business Venturing*, *18*(2), 165–187.

Huang, L., & Knight, A. P. (2017). Resources and relationships in entrepreneurship: An exchange theory of the development and effects of the entrepreneur-investor relationship. *Academy of Management Review*, *42*(1), 80–102.

Jack, S., Dodd, S. D., & Anderson, A. R. (2008). Change and the development of entrepreneurial networks over time: A processual perspective. *Entrepreneurship and Regional Development*, *20*(2), 125–159.

Johannisson, B. (1988). Business formation: A network approach. *Scandinavian Journal of Management*, *4*(3-4), 83–99.

Kacperczyk, A. J. (2013). Social influence and entrepreneurship: The effect of university peers on entrepreneurial entry. *Organization Science*, *24*(3), 664–683.

Kim, P. H., & Aldrich, H. E. (2005). Social capital and entrepreneurship. *Foundations and Trends in Entrepreneurship*, *1*(2), 55–104.

Krackhardt, D. (1992). The strength of strong ties: The importance of philos in organizations. In N. Nohria & R. Eccles (Eds.), *Networks and Organizations* (pp. 216–239). Harvard Business School Press.

Krishnan, R., Cook, K. S., Kozhikode, R. K., & Schilke, O. (2020). An interaction ritual theory of social resource exchange: Evidence from a Silicon Valley accelerator. *Administrative Science Quarterly*, *66*(3), 659–710.

Kumar, N., Stern, L. W., & Anderson, J. C. (1993). Conducting interorganizational research using key informants. *Academy of Management Journal*, *36*(6), 1633–1651.

Lamine, W., Jack, S., Fayolle, A., & Chabaud, D. (2015). One step beyond? Towards a process view of social networks in entrepreneurship. *Entrepreneurship & Regional Development*, *27*(7-8), 413–429.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691–710.

Langley, A., & Abdallah, C. (2011). Templates and turns in qualitative studies of strategy and management. In *Building methodological bridges: Research methodology in strategy and management* (pp. 201–235). Emerald Group.

Larson, A. (1991). Partner networks: Leveraging external ties to improve entrepreneurial performance. *Journal of Business Venturing*, *6*(3), 173–188.

Levi, M. (1996). Social and unsocial capital: A review essay of Robert Putnam's Making Democracy Work. *Politics & Society*, *24*(1), 45–55.

Lin, N. (2001). *Social Capital: A Theory of social structure and action*. Cambridge University Press.

Lorenzoni, G., & Lipparini, A. (1999). The leveraging of interfirm relationships as a distinctive organizational capability: a longitudinal study. *Strategic Management Journal*, *20*(4), 317–338.

McMullen, J. S., & Dimov, D. (2013). Time and the entrepreneurial journey: The problems and promise of studying entrepreneurship as a process. *Journal of Management Studies*, *50*(8), 1481–1512.

McMullen, J. S., & Shepherd, D. A. (2006). Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. *Academy of Management Review*, *31*(1), 132–152.

Miles, M. B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Sage.

Miller, K. D. (2007). Risk and rationality in entrepreneurial processes. *Strategic Entrepreneurship Journal*, *1*(1-2), 57–74.

Nanda, R., & Sørensen, J. B. (2010). Workplace peers and entrepreneurship. *Management Science*, *56*(7), 1116–1126.

Nebus, J. (2006). Building collegial information networks: A theory of advice network generation. *Academy of Management Review*, *31*(3), 615–637.

Neergaard, H., & Ulhøi, J. P. (2007). *Handbook of qualitative research methods in entrepreneurship*. Edward Elgar Publishing.

Obstfeld, D. (2005). Social networks, the tertius iungens orientation, and involvement in innovation. *Administrative Science Quarterly*, *50*(1), 100–130.

Pachucki, M. A., & Breiger, R. L. (2010). Cultural holes: Beyond relationality in social networks and culture. *Annual Review of Sociology*, *36*, 205–224.

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Sage.

Pentland, B. T. (1999). Building process theory with narrative: From description to explanation. *Academy of Management Review*, *24*(4), 711–724.

Pittaway, L., Robertson, M., Munir, K., Denyer, D., & Neely, A. (2004). Networking and innovation: A systematic review of the evidence. *International Journal of Management Reviews*, *5*(3-4), 137–168.

Porter, C. M., & Woo, S. E. (2015). Untangling the networking phenomenon: A dynamic psychological perspective on how and why people network. *Journal of Management*, *41*(5), 1477–1500.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, *24*(1), 1–24.

Pratt, M. G. (2008). Fitting oval pegs into round holes: Tensions in evaluating and publishing qualitative research in top-tier North American journals. *Organizational Research Methods*, *11*(3), 481–509.

Putnam, R. D. (1993). The prosperous community: Social capital and public life. *The American Prospect*, *13*(4), 35–42.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, *6*(1), 65–78.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

Rost, K. (2011). The strength of strong ties in the creation of innovation. *Research policy*, *40*(4), 588–604.

Saunders, M., Lewis, P., & Thornhill, A. (2012). *Research methods for business students*. Pearson education.

Saxenian, A. (1994). *Regional networks: Industrial adaptation in Silicon Valley and route 128*. Harvard University Press.

Scott, J. (2011). Social network analysis: Developments, advances, and prospects. *Social Network Analysis and Mining*, *1*(1), 21–26.

Shepherd, D. A., Douglas, E. J., & Shanley, M. (2000). New venture survival: Ignorance, external shocks, and risk reduction strategies. *Journal of Business Venturing*, *15*(5-6), 393–410.

Shepherd, D. A., Williams, T. A., & Patzelt, H. (2015). Thinking about entrepreneurial decision making: Review and research agenda. *Journal of Management*, *41*(1), 11–46.

Soetanto, D. P., & Jack, S. L. (2013). Business incubators and the networks of technology-based firms. *The Journal of Technology Transfer*, *38*(4), 432–453.

Strauss, A. L., & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Sage Publications.

Strauss, A. L., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications.

Stuart, T. E., & Sorenson, O. (2007). Strategic networks and entrepreneurial ventures. *Strategic Entrepreneurship Journal*, *1*(3-4), 211–227.

Suddaby, R. (2006). From the editors: What Grounded Theory is not. *Academy of Management Journal*, *49*(4), 633–642.

Teece, D. J. (1992). Competition, cooperation, and innovation: Organizational arrangements for regimes of rapid technological progress. *Journal of economic behavior & organization*, *18*(1), 1–25.

Tracy, S. J. (2010). Qualitative quality: Eight "big-tent" criteria for excellent qualitative research. *Qualitative Inquiry*, *16*(10), 837–851.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, *42*, 35–67.

Uzzi, B. (1999). Embeddedness in the making of financial capital: How social relations and networks benefit firms seeking financing. *American Sociological Review*, *64*(4), 481–505.

Van de Ven, A. H. (1992). Suggestions for studying strategy process: A research note. *Strategic Management Journal*, *13*(1), 169–188.

Van de Ven, A. H., & Engleman, R. M. (2004). Event- and outcome-driven explanations of entrepreneurship. *Journal of Business Venturing*, *19*(3), 343–358.

Vissa, B. (2012). Agency in action: Entrepreneurs' networking style and initiation of economic exchange. *Organization Science*, *23*(2), 492–510.

Walker, G., Kogut, B., & Shan, W. (1997). Social capital, structural holes and the formation of an industry network. *Organization Science*, *8*(2), 109–125.

Welter, F., Baker, T., Audretsch, D. B., & Gartner, W. B. (2017). Everyday entrepreneurship: A call for entrepreneurship research to embrace entrepreneurial diversity. *Entrepreneurship Theory and Practice*, *41*(3), 311–321.

Woolcock, M. (1998). Social capital and economic development: Toward a theoretical synthesis and policy framework. *Theory and Society*, *27*(2), 151–208.

Yen, Y. F., Tseng, J. F., & Wang, H. K. (2015). The effect of internal social capital on knowledge sharing. *Knowledge Management Research & Practice*, *13*(2), 214–224.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Sage Publications.

Yin, R. K. (2003). Designing case studies. *Qualitative Research Methods*, *5*(14), 359–386.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed.). Sage Publications.

Junior Management Science 9(3) (2024) 1733-1768

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Developing and Maintaining a Strong Corporate Culture, While Coping With a Workforce Growing Significantly: A Qualitative Analysis on Corporate Culture Development of Fast-Growing Start-Ups

Anna Simon

Technical University of Munich

Abstract

The development and maintenance of a strong corporate culture are crucial for the success of start-ups, especially during their founding and growth phases. While previous research emphasizes the founders' role in shaping corporate culture, the growth phase of start-ups in this context has hardly been explored. This study, based on 16 semistructured interviews with founders and managers of rapidly growing start-ups, provides new insights into corporate culture development and maintenance. The qualitative study reveals that various instruments play a vital role in shaping corporate culture during the founding and growth phases. Specifically, the founders' role model function, the definition of corporate values, recruiting, and events significantly impact this process. Additionally, the study highlights that start-ups actively and consciously strive to develop and maintain their corporate culture as the company expands in size. Given the numerous identified instruments for corporate culture development in start-ups, this work offers new insights into a relatively unexplored environment, serving as a foundation for further research. Furthermore, the practical implementation of these identified instruments is demonstrated, contributing to the practical value of this study.

Keywords: cultural development; organizational culture; rapid growth; start-ups

1. Introduction

"Cultures are like precious and prized treasures when they are strong, healthy, and driving the right behaviors. They are among the greatest assets an organization can have. However, they are vulnerable assets that can be damaged or lost if leaders are not aware of their value and are not keeping watch over possible culture-changing practices, attitudes, threats, or events." - Warrick (2017, p. 5)

According to the German Federal Ministry of Labor and Social Affairs (BMAS), corporate culture represents a significant potential for the success and competitiveness of companies in Germany. Thus, corporate culture and the associated employee commitment (O'Reilly, 1989, p. 17) also have a significant influence on the financial success of a company (Hauser et al., 2008, p. 29). Since corporate culture is seen as a driver for sustainable performance and competitiveness of companies, the topic has been in the focus of economic and social considerations for several years (Hauser et al. 2008, p. 31; Sackmann 2017, p. 11). Generally, research has shown that corporate culture evolves through the personalities of organizational members over time (O'Reilly et al. 2014, p. 596; Schein 1983, p. 3) and the mutual experiences of learning, failures, and successes (O'Reilly 1989, p. 19; Schein 1983, pp. 15, 20). This is why the culture gives each company a distinctive character (Janićijević 2011, p. 74; Schneider et al. 2013, p. 380; Willcoxson and Millett 2000, p. 93).

But what is the relevance of corporate culture, in particular for start-ups?

To compete with established competitors, start-ups need to be faster, more flexible, more innovative, and more efficient (Grossmann & Slotosch, 2015, p. 242). Even though they operate in an environment characterized by high uncertainty (Ouimet and Zarutskie 2014, p. 2; Sauermann 2018, p. 5) and have limited resources at their disposal (Katila et

DOI: https://doi.org/10.5282/jums/v9i3pp1733-1768

al. 2012, p. 17; Reypens et al. 2020, p. 13), they have to retain the best employees, from whom above-average commitment is demanded. Achieving this requires an outstanding corporate culture and its communication (Grossmann & Slotosch, 2015, pp. 242-243). Cummings (2011) even sees the corporate culture as the "only sustainable competitive advantage" (p.1) of start-ups that is entirely in the founders' control, which is why it has to be promoted and further developed within the firm.

Since the foundation of corporate culture is laid in the early days of a company, the founders are considered the most influential architects of the company's culture (Schein 1983; Schein 2004, pp. 225-227). Through this influence, founders have a unique opportunity to shape and develop a corporate culture that reflects a set of values, beliefs, and principles that endorse and reinforce the company's business purpose and strategy (Picken 2017, p. 8; Schein 1983, pp. 5-6). A lack of attention to the elements of creating and fostering a positive, strong corporate culture leads to a culture creating itself (Warrick, 2017, p. 9) and possibly a dysfunctional one accelerating a company's failure (Picken, 2017, p. 3).

In academic research, mechanisms, and means by which entrepreneurs shape and influence corporate culture have been studied (O'Reilly et al., 2014; Schein, 2004; Zheng et al., 2009). Yet a holistic view from the founders' perspective and the practical implementation of such mechanisms in a start-up context has not yet been sufficiently explored. Prior research has either focused only on individual mechanisms or taken a different research approach to identify them.

As the company ages and grows in personnel, the founders usually become less of a personal force; the trend is away from a sense of community to a more bureaucratic organization with one or more management levels that potentially care less about the original assumptions and values of the founders. This development is often feared by first-generation employees (Schein, 1983, pp. 23-26). According to Sackmann (2017, p. 26), especially rapid company growth poses not only particular challenges for areas such as logistics and operations but also for corporate culture. Company growth goes hand in hand with the need to integrate new employees into the corporate culture. A particular challenge in such a context is maintaining the specific corporate culture that contributed to the rapid growth and, if necessary, adapting it to the company with its increasing size and age. Therefore, it is essential to sustainably pass on the core of the corporate culture to the many new employees.

Despite the importance of corporate culture for start-ups, it frequently occurs that corporate culture is neglected during rapid growth and deprioritized alongside other growth challenges (Schmitt, 2018). This also seemed to have been the case with the direct bank N26, a German hyper-growth start-up founded in 2013, counting 1500 employees today (N26, 2022). In the last few years, negative headlines surrounding the start-up have increasingly emerged (Gründerszene, 2021). In 2019, the start-up gained three million new customers and tripled its headcount to 1500. However, strong criticism was voiced regarding the start-up's corporate culture, which was also reflected in high staff turnover and the demand for a works council, which the company's management initially opposed (Gründerszene, 2021; Zacharakis, 2020). Although the company continues to be in an intense growth phase, requiring concomitant staff growth, the number of employees has decreased by more than 300 between 2020 and 2021 (Hunter, 2021). On employee review portals such as kununu.com, the largest employer rating platform in the German-speaking area, the start-up is also given a recommendation score of 2.7 on a scale of 1 to 5. The rating considers the assessment of corporate culture since it is listed as an evaluation criterion alongside diversity, work environment, and career and salary. This value is lower than the banking industry average of 3.5 . The contentment with the corporate culture amounts to 3.2 points (kununu, 2022a, 2022b).

Although corporate culture has been researched in various contexts over the past few decades, the start-up context, especially of fast-growing ones, has hardly been the focus of researchers, even though they have gained increasing attention in public in recent years (Crosby, 2018; Hoffman, 2017; McGregor & Doshi, 2015; Schmitt, 2018) and are considered the driving force for innovation and the engine of economic growth (Kollmann et al. 2021, p. 17; Minola et al. 2015, p. 5; Reypens et al. 2020, p. 3).

This work aims to fill this research gap by examining corporate culture in the context of fast-growing start-ups. In this area, a better understanding of how corporate culture is actively shaped and developed by founders and managers, especially in the growth stage. Furthermore, the derivation of concrete instruments, practices, and measures for founders and managers is necessary to preserve the corporate culture at its core in their fast-growing start-ups. From this, the following research question has been derived:

> How can a start-up's corporate culture be positively developed and maintained while coping with a workforce growing significantly?

To answer this research question and, more specifically, to understand the founders' perspectives on how their start-up's corporate culture is developed and maintained during rapid growth, semi-structured interviews were conducted in which sixteen founders and managers from fourteen fast-growing start-ups participated.

This thesis is structured as follows: The subsequent chapter provides an overview of the existing literature on corporate culture and puts it in the context of growth, followed by Schein's theoretical framework on how founders embed their beliefs, values, and assumptions. In Chapter 3, the methodological approach of the thesis is consecutively described, followed by the findings of the interviews with the founders and managers of the start-ups in Chapter 4. In the next Chapter, the results are discussed, theoretical and practical implications are derived, and the limitations of this work and further areas of research are presented, leading to a conclusion in the last Chapter.

2. Theoretical Background on Corporate Culture

To elaborate on how corporate culture can be developed and maintained in a company's day-to-day operations, it is first necessary to understand the term corporate culture. To this end, the following sections explain what corporate culture means and its significance for companies. It will then be shown what influence the stages of founding and growth of a company's life cycle have on corporate culture. Finally, twelve mechanisms are presented according to Schein, through which corporate culture can be embedded and reinforced.

2.1. The Concept of Corporate Culture

The construct of corporate culture is a common concept in both practice and theory. It has already been studied by a wide array of theoretical interests with significant differences in the conceptual perspectives and methods used, from which a broad conceptual landscape and numerous academic debates have emerged over the last few decades (cf., Deal and Kennedy 1982; Hofstede 1991; Kotter and Heskett 1992; Martin 2002; Schein 1992; Schultz 1995; Trice and Beyer 1993). However, no consensus has been reached on its definition "given the vast array of approaches for conceptualizing and understanding culture" (Ehrhart et al., 2014, p. 130) and its multilayered and multidimensional nature, which is why the concept of (corporate) culture "remains an elusive and fuzzy concept" (Sathe, 1983, p. 6) for scholars and practitioners. Therefore, the following sections attempt to identify the most common approaches to defining corporate culture and distinguish it from the related construct of corporate climate.

2.1.1. Delimitation of Corporate Culture from Corporate Climate

Corporate culture and *corporate climate* are often equated in everyday use or even used as substitutes in the context of social constructs in companies (Sackmann, 2017, p. 64). Corporate culture and corporate climate are, in fact, two different concepts that have different perspectives on organizational environments and cannot be substituted for each other (Denison 1996, p. 625; Glisson and James 2002, p. 625; Pettigrew 1990).

Similar to the case of corporate culture, there is an ongoing debate in academia about how to define *corporate climate* for a scientific investigation (Glisson and James 2002, pp. 767-768; Guion 1973, p. 121; Schneider et al. 2013, p. 362). One approach to defining corporate climate originates from James and colleagues (James 1982, p. 229; Jones and James 1979, p. 201): If there is agreement among "employees in a particular work unit (...)on their [individual] perceptions of the impact of their work environment, their shared perceptions can be aggregated to describe their [corporate] climate" (Glisson & James, 2002, p. 769). First coming into the focus of research in the 1960s and 70s, *corporate climate* has been studied primarily quantitatively with surveys, based on a psychological research approach. On the other hand, *corporate culture* dominated research on the human organizational environment in the early 1980s with a sociological, anthropological, and mainly qualitative approach. *Climate* researchers have usually placed more emphasis on the situational perceptions of organizational members regarding observable policies, practices, procedures, and behaviors (Denison 1996, p. 622; Schneider et al. 2013, p. 362). Having referred to a snapshot, corporate climate is temporary, subjective, and more likely to be exposed to direct manipulation. On the other hand, *corporate culture* is rooted in the shared and partly subconscious value system and corporate history. It is collectively held, which is why it is considered more stable and less easily manipulated (Denison, 1996, p. 644). Once the corporate culture is deeply anchored in the behavior of employees, it can only be changed with great effort (Bryman 1986, p. 52; Ehrhart et al. 2014, p. 131; Schein 2004, p. 36).

2.1.2. Definition of Corporate Culture

After the distinguishing characteristics between corporate culture and corporate climate have been pointed out in the previous section, the question remains open as to what exactly is meant by corporate culture.[1]

Due to the complexity of the multi-layered construct of corporate culture and different research approaches, this results in different definitions. O'Reilly and Chatman (1996) describe corporate culture as "a system of shared values that define what is important and norms that define appropriate attitudes and behaviors for organizational members how to feel and behave" (p. 160); Hofstede (1984) defines corporate culture as "the collective programming of the mind that distinguishes the members of one human group from another" (p. 25); Kobi and Wüthrich (1986, p. 23) describe corporate culture as the set of norms, values, and attitudes that shape the behavior of all employees and thus denotes the way a company approaches things and solves problems. Martin (2002) entitles corporate culture briefly as "how things are done around here" (p. 3). Edgar Schein provided one of the most frequently acknowledged definitions of corporate culture. He defines the construct as follows: corporate culture is "a pattern of shared basic assumptions learned by a group as it solved its problems of external adaptation and internal integration, which has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel about those problems" (Schein, 2004, p. 17).

Despite the widely differing definitions of the corporate culture concept, individual aspects can be identified that have a common denominator, regardless of the researchers' theoretical orientation. Common to most definitions are the

[1] In literature, the terms corporate culture and organizational culture are used interchangeably. While organizations encompass all forms of organizations, corporates focus on for-profit companies. Since investigated start-ups in this study also pursue the latter, the term corporate culture is used in this thesis.

constructs “values” and “norms” as well as aspects of the cognitive level such as “basic assumptions”, “beliefs”, or “way of thinking” (Kaschube 1993, p. 97; Staehle 1999, p. 498).

2.1.3. Levels of Corporate Culture

In literature, corporate culture is usually understood as a multilevel concept whose levels differ in how accessible and visible cultural information is (Ehrhart et al., 2014, p. 291). The levels are roughly differentiated between how deeply culture is embedded in the psychology of members of an organization (cognitive elements, e.g., assumptions, values, norms, and attitudes) and to what extent the facets of culture are observable, particularly to outsiders (symbolic elements, e.g., language, behavior, artifacts) (Janićijević, 2011, pp. 72-73).

Beyond disciplinary boundaries, Edgar Schein’s three-level model for classifying the concept of culture and the associated phenomena has widely found recognition and application. Schein (Schein, 2004, pp. 25-27), divides culture and its phenomena into three levels: *Artifacts, espoused values and beliefs, and underlying assumptions:*

(1) *Artifacts*, representing the outer layer, are objectively visible elements and tangible manifestations of a culture, but they are often ambiguous to interpret without a corporate context. It is the observable behavior of the people in a company or the visible results of their actions. Artifacts include, for example, communicative behavior, rituals, organizational structures, processes, and even more materialized elements such as dress code or the architecture and design of an office building. Even if outside observers are able to detect these artifacts of a company, conclusions of their deeper meaning cannot be drawn without insider knowledge.[2] The next level comprises (2) *espoused values and beliefs*, which is a declared set of values and norms that provide information about desired social behavior or how to conduct business in particular situations - what is right or wrong or appropriate or inappropriate. Those values and beliefs are not visible; however, they can be articulated. Yet, it may deviate from the actual lived reality of the company. The (3) *underlying assumptions* of organizational life correspond to culture’s deepest layer and essence. They are difficult to articulate, intangible, and often understood only from an insider perspective. They are so deeply anchored that they have a decisive influence on the organizational members’ perceptions, ways of thinking, and acting. Those assumptions are the “ultimate source of values and action” and are primarily “unconscious and taken-for-granted” (p. 26) and difficult to change (Schein, 2004, p. 26-35).

The three levels of corporate culture interact with each other. The artifacts and espoused values build on the basic assumptions of the company members and make them visible to the outside world. By making them visible, they, in turn, also reinforce the basic assumptions of the individuals (Gontard, 2002, p. 27).

To preserve the differentiated nature of this work, it needs to be mentioned that other definitional approaches, which consider culture as something consisting of multiple layers of accessibility and visibility, mainly do not distinguish between the layers (2) and (3) of the three-layer concept of Schein. They tend to divide culture into two main layers. Hofstede et al. (1990) distinguish between the invisible shared (1) *values* that correspond to the core of culture and (2) *practices* that are visible to the observer, which includes symbols, heroes, and rituals. Symbols are visuals, objects, words, or gestures with a specific meaning within a given culture. Heroes are real or imagined people who possess valued characteristics and thus function as cultural role models. Rituals are socially essential activities “carried out for their own sake” (p. 291). Trice and Beyer (1984) describe two primary components of corporate culture: (1) its *substance*, meaning the “networks of ideologies, norms, and values” and (2) its *forms*, i.e., the practices through which these “meanings are expressed, affirmed and communicated to its members” (p. 654).

Despite the numerous possible interpretations of cultural layers, a common theme among all these definitional approaches is that corporate culture is built on values and beliefs shared by its members (Denison, 1996, p. 624). Even if Schein used the expression of “underlying assumptions” to express the core of corporate culture, he remarked that by “assumptions”, he describes what the majority of other culture scholars refer to as basic values and beliefs (Schein, 2004, p. 25). Considering this, it is reasonable to draw on the term “values and beliefs” most frequently used in culture literature to describe the core of corporate culture (Sackmann, 2017, p. 77). A second theme throughout all definitions is that there is a visible part to culture, whether they are called artifacts, practices, or forms. Therefore, these two elements are seen as the basic layer concept of this thesis.

2.2. Relevance of a Strong Corporate Culture

Now that the general concept behind corporate culture has been explained in more detail, it remains to be clarified what a strong corporate culture means and why it is relevant to a business.

According to academic research, a strong corporate culture is defined as one whose underlying values and beliefs that define what is important are “widely shared and strongly held” by members of a company (O’Reilly & Chatman, 1996, p. 166). The more agreement and acceptance about values and beliefs prevail and permeate a company, the stronger a corporate culture can be considered to be (Eberhardt 2013, p. 10; Ehrhart et al. 2014, p. 173). Therefore, a corporate culture with more widely shared values and beliefs has a stronger influence on employee behavior and a more far-reaching impact (Sathe, 1983, pp. 12-13).[3]

[2] Schein cites the example of the pyramids, which were built by both the Egyptians and the Mayans, but whose meaning was different. While for some, it was a burial place, for others, it was both a burial place and a temple Schein (1992, p. 30).

[3] Values, norms, and practices that are not well understood, unclear, inconsistent, or not reinforced is referred to as having a weak culture. Because

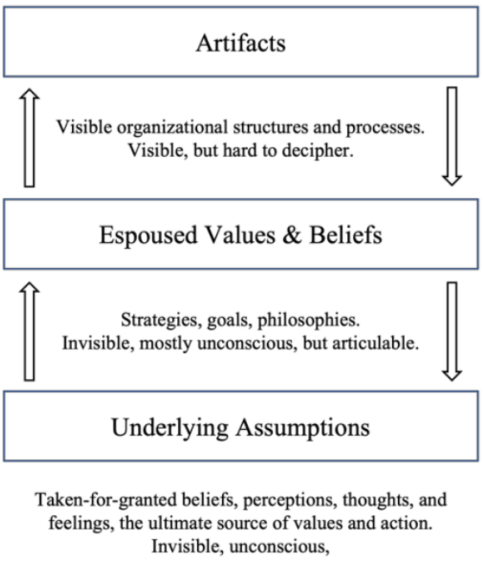


Figure 1: Levels of Culture according to Schein
(Source: Own illustration modified from Schein (1992, p. 30))

Sørensen (2002, pp. 6-7) summarizes the positive effects of a strong corporate culture that might lead to corporate effectiveness as follows:

A strong corporate culture has the critical consequence of increasing behavioral consistency among employees within a company. In this sense, corporate culture is an informal, social (1) *control mechanism*, influencing or even controlling employee behavior (O'Reilly, 1989, p. 11). In addition, it facilitates (2) *goal alignment* between the company and its members. It serves as a "vehicle for communicating and accomplishing organizational goals" (Ehrhart et al., 2014, p. 161). This means that employees know how to behave and how to deal with certain situations based on the common value system. If there is clarity about company goals and practices, employees tend to be more certain about the right course of action in unfamiliar situations and are able to respond appropriately. Aligning goals, in turn, also promotes (3) *coordination*, as there is less internal debate about the company's best interests (Crémer, 1993, pp. 15-17). Furthermore, strong cultures can increase employee (4) *motivation and commitment* if the "individual finds the values of the organization to be intrinsically rewarding and congruent with personal values" (O'Reilly, 1989, p. 18).

The fact that corporate culture has an influence on the above-mentioned aspects, Management literature implies that cultural strength enhances firm performance, i.e., the stronger the corporate culture, the more effectively a company operates (cf., Denison 1990; Denison and Mishra 1995; Kotter and Heskett 1992; Sørensen 2002; Waterman and Peters 1982).

there are ambiguous expectations and inconsistent practices throughout the organization, weak cultures have less influence on employees' behavior. Because of this, weak cultures typically exhibit inferior performance than strong cultures (Eberhardt 2013, p. 10; Warrick 2017, p. 5).

2.3. Linkage between Corporate Culture and Growth

Since the focus of this master's thesis is on start-ups that are in the growth phase, it is necessary that the following section embeds corporate culture in this context.

It is assumed that corporate culture evolves as the company develops, passing through different phases: Founding, Growth, Maturity, and Revival or Decline (Miller & Friesen, 1984, p. 1161). However, there are only a few studies that approach corporate culture from a dynamic perspective (cf., Hatch 1993; Schein 2004; Weeks and Galunic 2003; Zheng et al. 2009). According to Zheng et al. (2009, p. 158), who build their corporate culture evolution model on Miller and Friesen's (1984) conceptualization of the life cycle of an organization, the different phases of corporate development are accompanied by different corresponding cultural modes of action which is illustrated in Figure 2.

In the following sections, both the growth as well as the founding phase from Zheng and colleagues' model will be presented, as the latter is the relevant basis to reflect on how to develop but also to maintain corporate culture.

2.3.1. Founding Phase

During the founding phase, the primary concern of a company is the survival of the organization in the face of external turbulence, which is why attention is primarily paid to funding and marketing concerns; the establishment of structures and formalities is secondary (Kazanjian & Drazin, 1989, pp. 1489-1500). In this phase, companies tend to use cultural mechanisms that focus on keeping their members together rather than dealing with internal conflicts, along with transmitting the cultural assumptions and intangible values of the founding members to the employees and creating a collective identity instead of implementing those cultural values into practices (Zheng et al., 2009, p. 159).

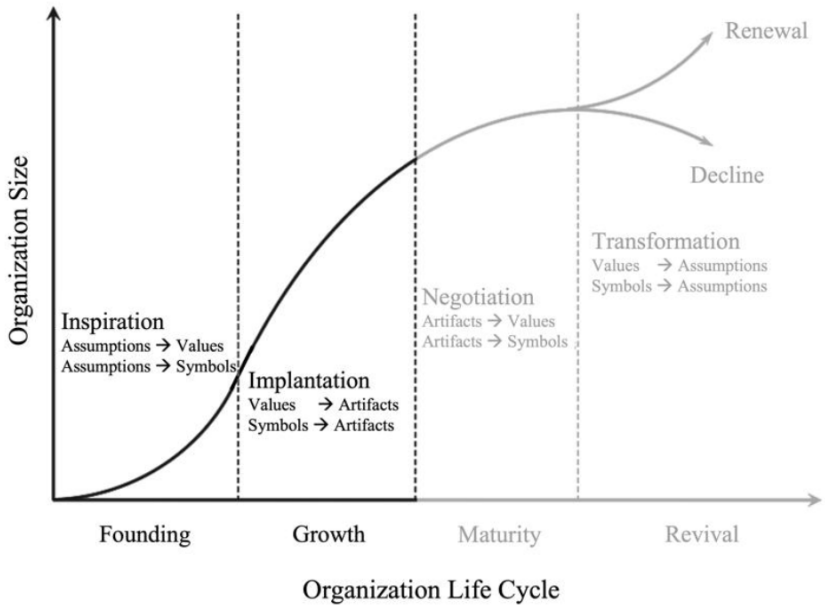


Figure 2: Organizational Life Cycle and Dominant Cultural Modes
(Source: Own illustration modified from Zheng et al. (2009, p. 159))

Zheng et al. (2009, pp. 159-160) postulate that in the founding phase the prevailing cultural mode is *inspiration*. By inspiration, they imply a cultural mode "that rallies organizational members through the leaders' shaping or aligning with members' values, beliefs, and aspirations" (p.160). This cultural mode is usually implemented through proactive manifestation and interpretation: the founders' fundamental assumptions are translated into values and communicated to employees through the company's vision and mission statements, inspirational stories about the company's founding, and giving the assumptions symbolic representations. Typically, this cultural mode works in a direction from the founding members to the employees.

2.3.2. Growth Phase

Once a company has successfully overcome its early survival challenges, it enters the phase of emerging growth (Zheng et al., 2009, p. 161). In this phase, the company faces several other issues, such as stabilizing production, meeting increasing demand, and formalizing structure (Dodge & Robbins, 1992, pp. 27-34). The search for new opportunities and their expansion represents the central concern at this stage (Jawahar & McLaughlin, 2001, p. 407). To sustain and facilitate growth, bureaucratic structures are developed, procedures are formalized, and authority is delegated to mid-level managers (Miller & Friesen, 1984, p. 1161).

In line with the "changes in strategy and structure, the primary cultural needs of the [company] shift, from identity formation to the consolidation of newly established cultural values" (Zheng et al., 2009, p. 161). The corporate culture, which was initially characterized by the charismatic value system of the founding members, usually evolves into a functionally based culture that aims to permanently maintain the corporate values, which in turn can align the members' behavior toward the company's mission (Wiener, 1988, pp. 537-539). Weeks and Galunic (2003, p. 1337) assume that the influence of founding members might still be strong but never guaranteed. They alone cannot ensure that specific values and assumptions continue to be effectively transmitted throughout the company during the further growth of the company. Instead, other cultural mechanisms have to be applied to attain that.

Zheng et al. (2009) refer to the dominant cultural mode in the growth phase as *implantation*: "Implantation refers to the process through which the cultural assumptions, values, and symbols that were diffused [in the founding phase] are implanted and embedded in a company in the form of organizational systems, structures, policies, rites and rituals, stories, and other tangible forms." (p. 161). To embed values into tangible forms and give artifacts symbolic meaning, implantation primarily entails proactive realization and symbolization. Therefore, this phase offers other key organizational members the opportunity to develop and maintain the corporate culture consciously and proactively (Zheng et al., 2009, pp. 161-162).

In this context, Schein (2004) uses the term *embedding* to describe how founding members and managers teach employees how to perceive, think, feel, and behave. For this, he identifies several mechanisms that founders, and managers can use to embed cultural values and beliefs. They are based on the assumption that, in addition to founders, managers are likely to have a greater influence on the development and shaping of corporate culture than other members of an organization (Eberhardt, 2013, p. 17). These instruments are explored in more detail in the following Chapter 2.4.

2.4. Mechanisms to embed and transmit Corporate Culture

As stated above, existing literature indicates that founders take a central role in influencing culture shaping in the founding phase, but more through their own direct and partly unconscious actions. As the size of the firm grows, there is a need for embedding cultural mechanisms to develop and maintain the company's culture intentionally.

The "teaching" process is the basic process of embedding values and beliefs in an organization. But the central question is: how do founders and key managers get the group to do things a certain way in the first place (Schein, 1983, pp. 14-15)? There are hardly any studies in the existing literature dealing with the concrete means founders or managers have at their disposal to communicate their team their innermost beliefs and values (Schneider et al., 2013, pp. 371372). Schein (1983, 2004) has perhaps been the most specific and detailed in describing how founders and key managers can embed, articulate and reinforce their values and beliefs. He proposes twelve mechanisms, distinguishing between mechanisms for embedding (primary mechanisms, see Chapter 2.4.1) and mechanisms for articulating and reinforcing values and beliefs (secondary mechanisms, see Chapter 2.4.2). For the mechanisms to be effective, they have to be consistent, i.e., words and actions must be congruent. If incontinence prevails, there is a risk that employees will misinterpret or reinterpret incidents, which can lead to a greater variety of assumptions throughout the company (O'Reilly and Chatman 1996, p. 21; Schein 2004, pp. 246-254).

This chapter describes the mechanisms, according to Schein (1983, 2004), that founders and managers can utilize to embed and reinforce assumptions, values, and beliefs (Schein, 2004, p. 270). Figure 3 gives an overview of Schein's twelve mechanisms.

2.4.1. Primary Mechanisms

The six primary embedding mechanisms depicted in Figure 3 are the major mechanisms for founders and managers to "teach their organizations how to perceive, think, feel, and behave" (Schein, 2004, p. 246) based on their own foundational beliefs and assumptions. Assuming consistency, the mechanisms operate simultaneously, interactively to a varying extent, and "reinforce each other to make the total message more potent than individual components" Schein (1983, p. 16). These mechanisms are described in more detail below to demonstrate how founders and managers can embed their assumptions through their actions.

What Leaders pay Attention to, measure, and control on a regular Basis

This mechanism is generally about aspects to which a founder systematically pays special (no) attention and thereby expresses what is (not) important to the founder personally. This can be reflected in what is measured, controlled, and rewarded and what founders react to emotionally, for example, when specific values have been violated. Through visible reactions from a founder, employees gradually adjust their behavior to what they believe they consider desirable behavior. Thus, the founder's basic assumptions are adopted over time if this behavior leads to the desired results (Schein, 2004, pp. 246-254).

Deliberate Role Modeling, Teaching, and Coaching

It is recognized in academia that the behavior of founders and managers, as perceived by employees, is a significant factor in shaping corporate culture (Bennis 1986, p. 64; Hofstede et al. 1990; Schein 1983; Trice and Beyer 1993; Tsui et al. 2006). Founders and managers are usually well aware that they act as role models through their own behavior, which helps to pass on their assumptions and values to employees. If a founder assumes that hierarchies should not play a crucial role in their company, this assumption can be expressed through dress, by the founder mingling with employees or situating themself with the rest of the workforce (Schein, 2004, p. 258).

Schein (2004) distinguishes between planned behavior in a staged environment and casual, informal behavior in everyday work situations. Staged environments are, for example, videos or speeches in which founders welcome new employees or outline their explicit philosophy. Casual role model behavior, such as dealing with a customer on the phone or with other employees, observed by employees and unconscious to the founder, is a more effective teaching mechanism than staged messages. If founders take on the role of a coach, instructing, correcting, or confirming their employees in their work, they simultaneously emphasize what is important to them and communicate their values more explicitly Schein (2004, pp. 258-259).

Reaction to critical Incidents and Corporate Crises

According to Schein, the way how founders and other managers deal with crises exposes critical underlying assumptions and creates new norms and values. Crises - the definition of which is a matter of perception - are particularly important for the creation and transmission of culture. The intense emotional engagement in such phases "increases the intensity of learning" (p.254) of values and assumptions. For example, if founders themselves or an employee make a wrong decision that costs the company a lot of money, dealing with such situations can reveal deep assumptions and provide opportunities where these assumptions form the basis for collective learning and thereby become solidified (Schein, 2004, pp. 254-255).

Recruiting, Selection, Promotion, and Excommunication

A very effective method of embedding assumptions and values is recruiting and selecting new employees who fit a particular "type of person" (Van Vianen, 2000, p. 145). This mechanism seems to be one of the more inconspicuous ones, as most founders and managers unconsciously recruit people

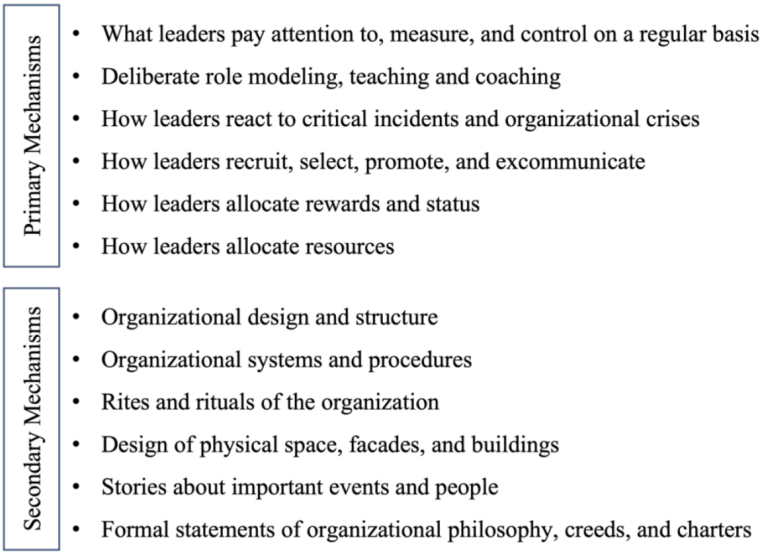


Figure 3: Mechanisms to embed beliefs, values, and assumptions
(Source: Own illustration modified from Schein (2004, p. 246))

for "person-culture fit" - people who are similar in the assumptions, values, and beliefs of existing members (O'Reilly et al., 1991, p. 492). By hiring people who already align with the founders' values, the founders directly reinforce the values they intend to embed. Thereby, the embedding process is facilitated (Schein, 2004, pp. 261-262). Chatman and Cha (2003, p. 26) recommend that when hiring new employees, greater weight should be given to a better fit with the corporate culture than existing skills as skills can be learned; a non-existent person-culture fit is more challenging to remedy.

Anchoring of values and beliefs also occurs through two other mechanisms, according to Schein (2004): internal promotion or firing and the criteria associated with them. Any change in employees and managers signals to other employees and managers what is desirable and not desirable behavior and what the company is looking for in its workforce (pp. 261-262). Thus, in the case of a promotion, it has to be questioned whether the person to be promoted represents the desired corporate culture (Sackmann, 2017, p. 268).

Allocation of Rewards and Status

Another powerful mechanism for founders and managers to communicate and manifest what they value is what they systematically reward (Drucker 1991; Kerr and Slocum Jr. 2005, p. 135). Employees learn what behavior is valued or undesirable internally through performance assessments, rewards, and conversations with their supervisors. As a result, not only the type of rewards are carriers of the corporate culture in the long run, but also the behavior they encourage. Therefore, if founders or managers want to ensure that their assumptions and values are learned and lived, care has to be taken that the respective rewards, and the status system align with those assumptions and values (Schein, 2004, p. 259). If, for example, a result orientation is to be rewarded, in which it is primarily the result achieved that is of importance and not the time spent at the workplace, this should also be reflected in a correspondingly performance-oriented compensation system (Sackmann, 2017, p. 276).

Allocation of Resources

Another mechanism that reveals the assumptions and beliefs of founders is the use, distribution, and approval of budgets, as these exert profound influence on the choice of goals and, consequently, the means to achieve them. At the same time, such beliefs serve as decision criteria for what corporate resources should be used for and constrain decision-making by narrowing the perception of alternatives. But it is not just about what money is spent on, but how. For example, the belief in a particular system, such as a bottom-up entrepreneurial system, can be clearly illustrated through a resource allocation process: engineers and managers below managers are encouraged to create business plans and budgets, so they would not be dictated from above but only approved by managers (Schein, 2004, p. 257).

2.4.2. Secondary Mechanisms

Unlike the primary mechanisms, the secondary ones, such as organizational structure, architecture, rituals, and formal statements, are focusing on reinforcing corporate culture. They only work when consistent with the primary mechanisms described earlier. What is learned informally at the beginning is now formalized (Schein, 2004, pp. 262-263).[4]

[4] Zheng et al. (2009, p. 168) refer to the mechanisms that transform values into artifacts and the symbolization of artifacts as "*implantation*," which - as mentioned earlier - is especially important in the growth phase of a company.

Organizational Design and Structure

Leading management academics such as Mintzberg (1979) have already shown how organizational structure and design can have a positive internal impact on members' thinking and behavior. The organization's design and structure determine the responsibilities within the company and regulate the division of labor, the formal information and communication channels, flow of authority and influence, and the form of cooperation. The founders' "deeply held assumptions about the task, the means to accomplish it, the nature of people, and the right kind of relationships to foster among people" (p. 264) can thus be embodied in the initial design of the organization and, if necessary, in regular restructuring. For instance, a decentralized organization is built if the founders assume that their employees are the experts and should make decisions themselves in their area of responsibility. It is designed to shift authority downward as much as possible (Schein, 2004, pp. 263-264).

Organizational Systems and Procedures

Routines, procedures, and other recurring tasks not only guarantee effective work, but also provide structure and predictability to a company operating in a vague, ambiguous organizational world. The systems and procedures thus reduce ambiguity and anxiety among corporate members. Founders and managers can strengthen their assumptions by establishing these systems and "daily, weekly, monthly, quarterly, and annual cycles of routines" (p. 265) around them. This clarifies what they consider essential and what employees need to pay attention to, i.e., this mechanism formalizes the process of "systematically paying attention" (p. 265). An example of this mechanism is a founder who reinforces his belief that truth can be achieved through debate by creating and participating in various committees (Schein, 2004, pp. 264-265).

Rites and Rituals of the Organization

Several cultural scholars regard organizational rites, rituals, and ceremonials as central to decoding and communicating cultural assumptions and maintaining established value systems (Trice and Beyer 1984, p. 656; Wiener 1988, p. 543). They combine diverse forms of cultural expression[5] into coherent cultural events with clearly delineated beginnings and endings. A classic integration rite that organizations frequently use is Company Christmas parties. They create, promote, and revive shared feelings that connect members and bind them to a social system (Trice & Beyer, 1984, pp. 654, 657). Another example is frequent off-site meetings with their own designations, locations and informal procedures (Schein, 2004, p. 267).

[5] E.g., language, gestures, artifacts, symbols, settings, or ritualized behaviors.

Design of Physical Space, Facades, and Buildings

All visible features of the company encountered by customers, incoming employees, and visitors are included under the mechanism of physical design. Messages from founders can be derived from the physical environment, such as in the case of structures and procedures, if the physical design is explicitly controlled (Schein 2004, p. 267; Steele 1973). Such physical artifacts reflect and reinforce underlying assumptions about how work gets done and how relationships, interactions, and communication flows should be managed. Research shows that the physical environment can influence the thinking and actions of managers and employees and should therefore be considered accordingly regarding the desired type of corporate culture. If founders have a clear philosophy and style, these can be embodied in the visible manifestations of their company. For example, an open office concept can express assumptions about equality, honest and direct communication, and the importance of relationships (Schein, 2004, p. 267-268).

Stories about important Events and People

As a young company and its group develop and build a history, part of that history transitions into stories about events and leadership (Allan et al., 2002; Martin & Powers, 1983). These stories not only reinforce the assumptions but are also taught to newcomers (Schein, 2004, p. 268).

Yet, it is often the case that the messages in the stories are highly distilled or even ambiguous and challenging to control. Thus, drawing correct conclusions from these stories is usually not possible without contextual knowledge. If values have first been anchored by primary mechanisms, these stories can serve to deepen and concretize that understanding of them. As with the mechanisms above, stories should not be used as the sole means of doing so (Schein, 2004, p. 268-269).

Formal statements of Organizational Philosophy, Creeds, and Charters

Another mechanism of articulation and reinforcement of cultural elements Schein (2004) mentions are formal statements, "the attempt by the founders for managers to state explicitly what their values or assumptions are" (p. 269). Usually, these statements reveal and highlight only a small part of the assumptions held in the group, which can be articulated formally as well. They emphasize specific values from the founding team's point of view that should be followed in the organization and not forgotten. Amazon's twelve Leadership Principles provide an example of this (Amazon, 2019).

However, further literature points out that it is not only communication of formal statements from the management side that contributes to fostering an existing corporate culture. Culture is also communicated through good interpersonal relationships and teamwork (Willcoxson & Millett, 2000, p. 97). Furthermore, choosing appropriate commu-

nication systems and language to communicate in a way that is consistent with and regularly reinforces the corporate culture is another cultural lever (Men et al., 2018, p. 1).

In summary, primary mechanisms allow founders and managers to embed their assumptions and core values into the everyday life of their company if they are all compatible with each other. Secondary mechanisms are two-folded. Early in the growth phase of a company, messages conveyed through secondary mechanisms are "less powerful, more ambiguous, and more difficult to control" (p. 270). However, as the company matures and stabilizes, the secondary mechanisms become primary maintenance mechanisms (Schein, 2004, pp. 270-271).

3. Methodology

An initial review of the literature on corporate culture, its development from the founding phase to the growth phase and possible embedding instruments, and the identification of a research gap in the start-up context led to the derivation of the following "well-specified, if rather general, research question" (Gioia et al., 2013, p. 19):

> *How can the corporate culture of a start-up be positively developed and maintained while coping with a workforce growing significantly?*

The methodology outlined below is intended to help fill this research gap by expanding knowledge on the above topic and answering the research question.

In the following chapters, the chosen research design is first explained, followed by the sampling strategy, explaining why and how the research samples were selected. Next, the data collection method is presented, concluding with a transparent description of the data analysis procedure.

3.1. Research Design

This thesis's primary focus is to better understand how start-ups manage to develop and maintain their corporate culture while growing significantly in personnel. Given the limited academic background on corporate culture in fast-growing start-ups and how they deal with the topic, this study employs an exploratory *qualitative study design* to address and generate further insights on the research question (Janićijević 2011, p. 83; Mayring 2020, pp. 11-12).

The qualitative research approach is recommended for investigating insufficiently understood phenomena, as it is with corporate culture in the start-up context (Strauss & Corbin, 1998, p. 11). Compared to quantitative research, qualitative interviews often provide a deeper understanding of social phenomena such as individuals' experiences, perceptions, and beliefs (Neergaard & Leitch, 2015, p. 4). Thus, the chosen approach allows exploring corporate culture from an "insider perspective" expressed "in the language of the people and, thus, the firm" (Ehrhart et al., 2014, p. 142) and to describe the way the study participants understand, define, act, and manage their everyday situations in their particular organizational environment (Miles et al., 2018, p. 25). Furthermore, this research design aims not to determine a quantitative specification of named aspects; instead, the discovery of these is in the focus, which are represented by a diverse group of people (Witt, 2001, pp. 5-7).

For this purpose, sixteen semi-structured interviews were conducted with founders and managers of fourteen fast-growing start-ups to reconstruct and evaluate their past cultural development measures and identify instruments that have contributed to maintaining their corporate culture.

3.2. Sampling

Multiple cases of fast-growing German start-ups were sampled to investigate the research question. Start-ups are defined in this thesis according to the characteristics of Kollmann et al. (2021, p. 12): "Start-ups are younger than ten years, have a planned employee/revenue growth and/or are (highly) innovative in their products, services, business models and/or technologies".

To capture an in-depth understanding of the corporate culture of the respective start-ups, interview partners were selected with regard to their knowledge of the start-up's culture, also called "knowledgeable agents" (Gioia et al., 2013, p. 17), as they are the ones "that are especially knowledgeable about or experienced with a phenomenon of interest" (Palinkas et al., 2013, p. 534). This primarily includes founders who had a decisive role in shaping and influencing the corporate culture from the beginning (Schein, 1983), followed by high-level managers (C-Level) or employees who are among the company's first employees and are dedicated to the topic of the corporate culture. Because of the different positions within the company, they might view corporate culture in slightly divergent ways. However, the critical factor for this thesis was that each could report on how culture was managed within their start-up.

The samples were selected in such a way that wide-ranging information could be obtained regarding the research question while still preserving a research focus. Thus, a combination of theoretical and purposeful sampling was chosen as the sampling strategy.

The sample was iteratively selected according to the *theoretical sampling* (Glaser & Strauss, 1967). Since the primary research focus is to capture instruments that helped start-ups develop and maintain their corporate culture, heterogeneous cases are considered during the sampling process "for the likelihood that they will offer theoretical insight" (Eisenhardt & Graebner, 2007, p. 27). This is intended to cover the subject area as broadly as possible, gather important common patterns across cases, and thus, increase the robustness of the studies' findings (Palinkas et al., 2013, p. 535).

Additionally, *purposeful sampling* was used to identify relevant and information-rich cases that express the research subject of interest in-depth (Patton 2002, p. 273; Yin 2009, p. 18).

The following criteria were considered when selecting the respective start-ups:

Firstly, the fulfillment of the provided definition of start-ups, according to Kollmann et al. (2021, p. 12). Secondly, registration of the company headquarters in Germany. This is because corporate cultures emerge in broader cultural contexts such as national or ethical groupings (Willcoxson & Millett, 2000, p. 92). Thirdly, employment of at least 100 people. This is because, at this size, employees typically no longer know each other by name, which is why further managerial actions may be required regarding corporate culture (Crosby, 2018; Lourenco, 2021; Valencia, 2019). Fourth, a personnel growth rate of at least 20% over the last two years. According to the definition of OECD, enterprises are considered high-growth enterprises if they show an average annual growth of more than 20% over three years, whereby growth can be measured not only by turnover but also by the number of employees (Eurostat-OECD, 2007, p. 61). Since the professional social network LinkedIn was primarily used as a source for the personnel growth rates, but only the growth rate over the last two years can be viewed there, the criteria period was consequently limited to two years. Lastly, the corporate culture rating on kununu.com is at least 4.0 out of 5.0. Kunnu.com is Europe's leading platform for employer ratings and information on salary and corporate culture (kununu, 2022b). This criterion is intended to ensure more objectivity in assessing the "strength or attractiveness" of a start-up's corporate culture.

Initial research identified approximately 58 start-ups that met those predefined selection criteria in Germany, whereby no claim is made to completeness. From this initial funnel, fourteen start-ups were used as the data basis.

Geographically, the start-ups studied are located in Bavaria or Berlin. Most of the start-ups are founded in Munich. This is primarily due to the identification procedure used for the interview partners, based on the author's entrepreneurial network in Munich. Some contacts to interviewed representatives of start-ups could already be established via the network, allowing them to be contacted directly via Slack, LinkedIn, or email. In addition, some other interview partners who could not be reached through the network were also directly contacted via LinkedIn.

As there are no clear guidelines or specifications regarding the ideal number of interview partners (Guest et al., 2006, p. 59), interviews were conducted until the state of theoretical saturation occurred, which is described as the state at which no significant new information is gained through further interviews (Edwards & Holland, 2013, p. 65). In conducting this research, saturation occurred after the fourteenth start-up, as the marginal relevant information significantly decreased after this point.

3.3. Data Collection

Once the initial research question and sampling strategy are determined, a direction before and during fieldwork for how the information will be collected has to be determined (Witt, 2001, pp. 5-6).

As the objective of this study is to find out how start-ups manage to develop and maintain their corporate culture in a rapid growth phase, the interview data outlined in Chapter 3.2 is used as the primary source of data "to obtain both retrospective and real-time accounts by those people experiencing the phenomenon of theoretical interest" (Gioia et al., 2013, p. 19).

In addition, to the interview data, secondary data were collected before and after the interviews to expand case knowledge and enable data triangulation (Eisenhardt 1989, pp. 537-538; Yin 2018, p. 15). The data triangulation is intended to capture initial information about the corporate cultures of the respective start-ups, if possible, facilitating higher quality conversations or validating findings from the interviews. These include company websites, LinkedIn posts targeting the corporate culture theme, and past online interviews with the founders in written form.

Interview Design

The qualitative study data were mainly collected through semi-structured interviews that captured personal experiences, perceptions, self-reflection, and the explanations behind them regarding a specific problem (Adams 2015, p. 496; Witzel 2000, p. 1). The approach not only allows more flexibility to the interviewee's narrative mode but also to "follow wherever the informants lead us in the investigation of our guiding research question" (Gioia et al., 2013, p. 20). Besides, corresponding follow-up questions can be asked, which might yield unexpected results (Rubin & Rubin, 2011, p. 158). At the same time, sufficient comparability between the interviewees is ensured by predefined topic areas covered in the interview (Adams 2015, p. 493; Edwards and Holland 2013, p. 29).

For this purpose, an interview guide was created as a data collection tool, which can be divided into six thematic sections: A short introduction, which covered the purpose and goal of the study, set the initial atmosphere for the interview. Furthermore, the structure of the interview, anonymity of the data, the relevance of answering all questions reflectively and openly, and the reference to one's perception and experiences regarding their start-up were pointed out. To counteract another potential inhibition, it was stated that there were no right or wrong answers in the cultural context (Ehrhart et al., 2014, p. 153).

Regarding the narrative mode, open-ended and more generative questions were used in the following section, covering the topics of corporate culture more broadly (Strauss, 1987). This allowed the author to find out what the interviewees understood by corporate culture and which aspects they considered part of it, as approaches to culture management in organizations depend on the interviewee's conception of corporate culture (Willcoxson & Millett, 2000, p. 92). The focus was then directed in the fourth section to the past to find out how the founders approached the topic of corporate culture in the initial phase and which aspects have shaped it in a trendsetting manner. The fifth section dealt specifically with the growth phase and which instruments or measures they make use of, especially in scaling, to maintain the start-

up's culture. In the concluding sixth part, the interviewee was deliberately given time to reflect on any thoughts that remain open before the interviewer ends the interview with a word of thanks.

Particularly after the first interviews, the questions and possible improvements for the following interviews were reflected upon, which is why the interview guideline was iteratively adapted in the course of the research process according to the information obtained (Gioia et al., 2013, p. 19). The interview guide can be taken from Appendix A.

Interview Setting

The interviews were conducted remotely between November 2021 and early January 2022, primarily using the telecommunication software Zoom. The remote interview setting, mainly entailed by the Covid-19 pandemic and the resulting contact restrictions, allowed for a more diverse sample, as the location of the interviewees was not a restraining factor. All interviews were conducted in German, the native language of the interviewees, to prevent potential language barriers (Welch & Piekkari, 2006, pp. 428-429). To report findings, only representative quotes were translated. Furthermore, all interviewees agreed to an audio and video recording under the condition of anonymity. Moreover, by emphasizing confidentiality and anonymity of the interviews at an early stage, it was possible to reduce the uncertainty and skepticism of some interviewees (Huber & Power, 1985, p. 176).

To determine the demographic data of the interviewee as well as facts about the start-up, a short questionnaire was sent to the interviewees beforehand. With the help of this questionnaire, it was possible to classify later and evaluate what was said and intended to facilitate the entry into the conversation (Witzel, 2000, pp. 3-4). The short questionnaire can be taken from Appendix A. Thus, the time involved in an interview could be reduced to the interview time itself and most essential questions, allowing even busy founders and managers to be recruited for an interview.

After the first two interviews, it became apparent that the topic of corporate culture might seem somewhat abstract to the interviewees and that reflection on past years might also require some time for consideration. Subsequently, some guiding questions were sent to the study participants in advance. This was intended to get the interviewees into the right frame of mind for the topic and set the basis for answering the questions in a more targeted manner. Furthermore, situational questions were additionally asked to capture concrete examples from everyday business life and thereby gain a deeper understanding of the instruments and practices used in the start-ups to develop and maintain culture.

Interview Data - Overview

A total of sixteen semi-structured interviews within fourteen start-ups were conducted that met the selection criteria from Chapter 3.2. In two start-ups, two people agreed to participate in the study.

The interviews varied between 28 and 75 minutes, with an average duration of 44 minutes, whereby the researcher's introductory words are not included in the recording but only started from the time of questioning. In one case, questions were answered in writing as a supplement to a previously conducted interview to gather some additional information in the case of company C10 (cf. Eisenhardt 1989, p. 539).

In nine of fourteen start-ups, a person from the founding team could be gathered for an interview. In two cases, the interviewees are assigned to the C-level, and in three others to a management position. Overall, the sample is dominated by male interviewees with a respective share of 69 percent. Table 1 gives an overview of the interviews conducted and the corresponding metrics. To maintain anonymity, the start-ups and interview participants were given designations. They were numbered consecutively according to Table 1; if the interviewee was a founder, an F was prefixed; if it was a manager, an M precedes the numbering; start-ups received the abbreviation C for company.

The interviews resulted in 658 minutes of recording and 261 pages of transcript. Filler words and potential grammatical errors are documented in the transcripts to preserve the integrity of the conversation (Döring & Bortz, 2016, p. 312). The transcripts of the conversations are attached in Appendix C.

3.4. Data Analysis

For the analysis and evaluation of the collected data, the qualitative content analysis according to Mayring (2015) was conducted with the support of the software MAXQDA in software version 22.2.0, which is commonly used in qualitative content analyses (Kuckartz & Rädiker, 2010, p. 734)

The procedure for qualitative content analysis generally consists of two steps:

In a first step, categories inductively obtained from the material or deductively theory-based are assigned to individual text passages based on coding rules. In a second step, it is examined whether specific categories can be assigned to further text passages. Even though content analysis rules accompany the process, it is a qualitative-interpretative process, which is why latent meaning can also be captured (Mayring & Fenzl, 2019, p. 634).

Since there are only limited findings on the research topic of this study in the context of fast-growing start-ups, the procedure of inductive category formation with the help of in vivo and descriptive coding approaches was chosen. Within the framework of qualitative content analysis, according to Mayring, inductive category formation starts from the text material, i.e., the categories are derived directly from the material (Mayring, 2015, pp. 85-86). Within Grounded Theory, this procedure is called "open coding" (Glaser & Strauss, 1967).

As previously indicated, inductive category formation is a rule-based procedure, which the researcher of this thesis followed. The process is illustrated in Figure 4.

Table 1: Overview Sample and Interviews (Source: Own illustration)

# Start-up	# Participant	Role	Gender	Founding Year	Company City	Company Industry	# of Employees	# of Employees (July)	Growth Rate within 2 years	Kununu Rating	Funding Phase	Date of Recording	Duration in minutes
C1	F1	Co-Founder: Software & Product	m	2019	Munich	IT & Services	> 100	> 150	888%	4,9	Seed	16.11.21	32
C2	F2	Co-Founder: COO	m	2019	Munich	Automotive	> 170	> 300	> 1000%	4,8	Series A, raising B	19.11.21	31
C3	F3	Founder: previous CEO	m	2014	Munich	Industrial Automation	> 230	> 270	20%	4,2	Post Series B	24.11.21	62
C4	M4	Head of Strategy & Business Development	w	2019	Berlin	IT & Services	> 200	> 500	> 1000%	4,3	Series A, raising B	30.11.21	34
C5	F5	Co-Founder: CTO, Managing Director	m	2017	Munich	E-Learning	> 500	> 700	> 1000%	4,8	Series A	30.11.21	47
C6	M6	Chief of Staff	m	2018	Ottobrunn	Aerospace	> 220	> 250	513%	4,8	Series B	02.12.21	57
C7	F7	Co-Founder: CTO	m	2013	Munich	IT & Services	> 240	> 300	32%	4,5	Series C, Late VC	03.12.21	48
C8	M8	Head of Organization & People	w	2016	Munich	Automotive	> 245	> 350	138%	4,4	IPO	06.12.21	48
C9	M9	Academy Manager	w	2016	Munich	Environmental Services (B2B)	> 300	> 340	69%	4,9	n.a.	08.12.21	75
C10	F10; M10	Founder: CEO; First Employee: Team Lead HR	m; w	2018	Munich	PropTech	> 100	> 150	173%	4,9	Series A, raising B	14.12.21 17.02.22	30; n.a.
C11	M11	COO	m	2014	Munich	IT & Services	> 330	> 400	55%	4,6	Series D	14.12.21	32
C12	F12	Founder: Managing Director	m	2017	Berlin	Healthcare/ HealthTech	> 300	> 300	101%	4,1	Post Series B	15.12.21	53
C13	M13	CPO	w	2017	Munich	IT-Security /SaaS	> 180	> 200	106%	4,5	Series A, raising B	21.12.21	28
C14	F14a; F14b	Co-Founder: CTO; Co-Founder: CMO	m; m	2018	Munich	Food & Beverages	> 245	> 270	528%	4,6	Series B Late VC	7.1.21; 28.1.22	39 42

All data points except for the *# of Employees (July)* column were collected at the time of the interviews conducted. This column is intended to show a trend for the further growth of the start-ups and was collected at the end of July.

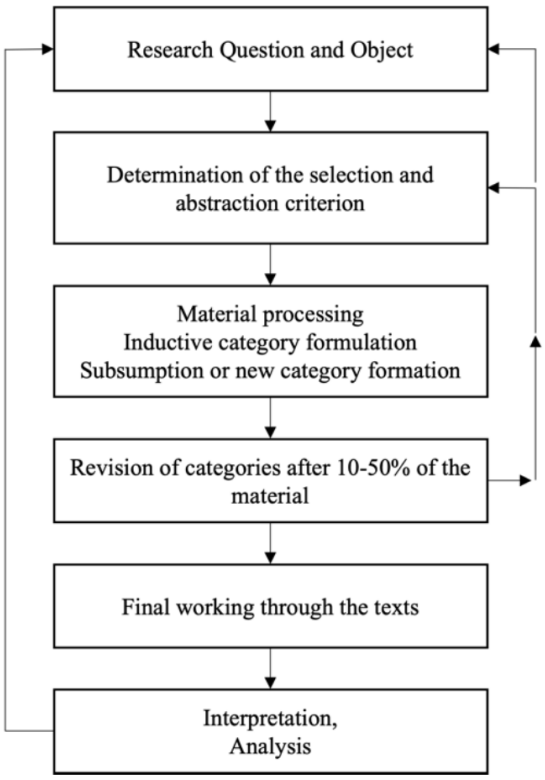


Figure 4: Process model of inductive category formation
(Source: Own illustration modified from Mayring (2015, p. 86))

The initial research question and thus factors that have contributed to the development and maintenance of corporate culture served as the selection criterion, and the interview transcripts were used as the starting point for the category definition. The level of abstraction was initially kept very concrete to conclude the specific to general patterns in the text at a later stage. Then the first text passages were systematically worked through, and the first categories were formed. In the further course, it was decided whether the following text passages fall under an already created category (subsumption) or a new category had to be formed. As soon as a large part of the material had been worked through, the revision started, and the system of categories developed so far was iterated. Finally, all the material was revised, with the categories formed up to that point.

The result of the inductive process is a system of categories linked to concrete text passages. Subsequently, main categories were formed. This step was also carried out, taking into account theoretical considerations according to Schein from Chapter 2.4 (deductive). However, as can be seen from the category system, new main categories were also formed, or existing ones were adapted based on the results that were not apparent in Schein's mechanisms. The entire system of categories is attached to Appendix B.

4. Findings

This study focuses on answering how start-ups manage to develop and maintain their corporate culture, even when the company is growing rapidly in terms of personnel. Based on the research methodology presented earlier, answers to the defined research question could be derived.

A wide range of instruments emerged from qualitative content analysis, 30 in total, which have contributed to corporate culture development and maintenance in the start-ups investigated. In the following, the term "instruments" refers to corporate guidelines, practices, measures, structures, and processes that have been utilized.

The results on the respective instruments, their meaning from the interviewees' point of view for their corporate culture, and concrete examples of use are presented below.

Since the interviewees repeatedly referenced the same instruments during the interview, the individual instruments are discussed in terms of the ten main categories formed, which are roughly based on the order of Schein's mechanisms rather than the thematic sections of the interview guide. As far as a statement about the relevance in the different life phases of the start-up was made, however, these are mentioned.

The resulting data structure is shown in Figure 5 and will be presented according to this order in the following chapters in detail.

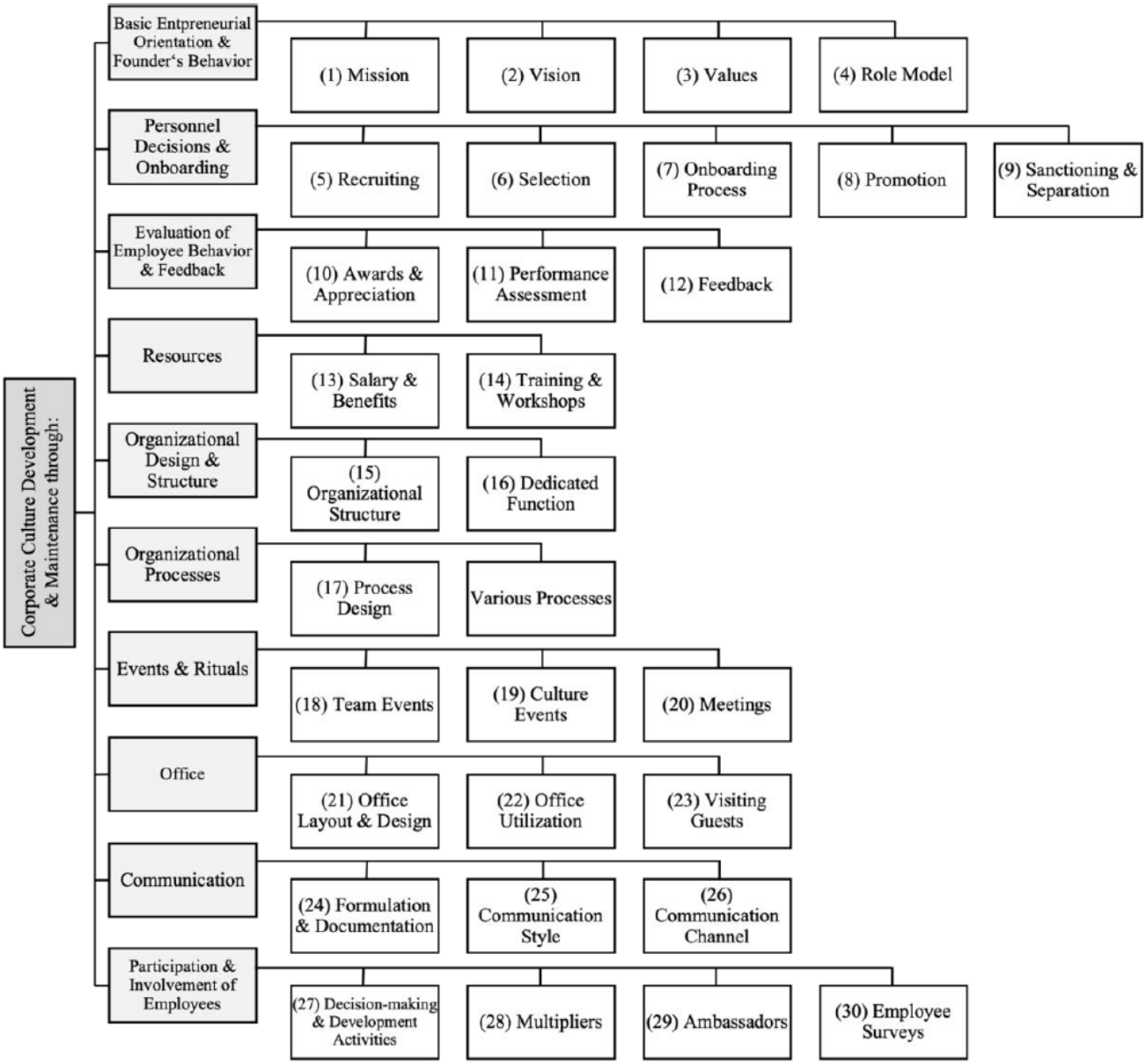


Figure 5: Instruments for Corporate Culture Development Maintenance - Overview (Source: Own illustration)

4.1. Basic Entrepreneurial Orientation and Founders' Behavior

This chapter deals with the mission, vision, and corporate values of the start-ups as a basic entrepreneurial orientation, as well as the behavior of the founders in their function as role models for their employees.

Some of the interviewees described the start-up's (1) mission, which provides "the fundamental reason why an organization exists" (Pearce & David, 1987, p. 109), i.e., the corporate purpose, as a guiding component and carrier of their corporate culture. The reason, they argued, is that their missions already imply a value system:

> "A lot of the founder's basic assumptions that are also explicit are all around our mission and the SDGs, and I think there's such a general basic connection, and it's not questioned because most people come to [C9] because the mission was the deciding reason there, and the basic assumptions of our founders that attract people (...) which is why along with these basic assumptions the discourse is still happening. Otherwise, we would have lost our mission. I think that founding mission and vision was just such a strong pull factor [for our culture] that they (...) are still equally important." (M9)

Since the mission includes a concrete corporate purpose for which the members of a company strive, a certain basic bond is already created between them (M8; M9; M13).

A similar role, but less frequently addressed, is assigned

to the start-up's (2) vision by envisioning the kind of environment a company would like to "create within a broad time horizon and the underlying conditions" for achieving that vision (El-Namaki, 1992, p. 25):

> "Our decisions are also very much vision-driven. (...) And that is also the unifying factor, that we all know very clearly what our goal is. And I think that is something that is also very clear to the new joiners from day one, which we also always get as feedback: Yes, we have a shared journey; we have a shared vision that we are working towards." (F12)

One study participant indicated that to build the foundation for a shared understanding of the vision among new employees, talking about the vision is part of the onboarding process (M11). Furthermore, another start-up introduced a "vision reminder" in a weekly meeting to anchor the vision with the existing team. The founders context-specifically address the vision and talk about it there (M8).

In addition, the vision and mission are partly seen as a key reason why people (would like to) work for the start-up in the first place (F5; M8; M9; M13).

Many of the interviewees referred in their definition of corporate culture to a value system that an organization is based on and with which the members of an organization informally agree (F2; M4; F5; M6; F7; M8; M9; M11). Even when they describe their own corporate culture, they usually refer to the start-up's (3) values (F3; F4; F7; M11; F12; M13). All of the start-ups studied had already been engaged in articulating corporate values, i.e., values, beliefs, and principles captured by verbal expressions, in part to express and communicate their implicit assumptions about the way the start-up should operate. Hence, some founders refer to these values as "guiding principles" (F7), "fundamentals" (M13), or "maxims for action" (F10) during the interviews.

For many start-ups, those values function as an important building block for initially setting the desired direction of the corporate culture and as a benchmark for establishing shared behavioral guidelines (F1; F2; M4; M8; F12; M13; F14a). A founder's following statement shows that the formulated corporate values serve internally as decision guidelines and are internally repeatedly referred to as a reference:

> "Every project and goal must always be evaluated: is it aligned with our values? We also pay much attention to this in the formulation of goals, but also in the projects themselves, so that we say: to which corporate value does the project contribute?" (F14a)

Other managers in the sample attributed importance to the corporate values in challenging situations where a decision must be made, or an appropriate solution be found. Here, again, they are used as a reference point (M6; M9).

Except for C5, all of the start-ups have formalized these values in writing (more on this in Chapter 4.9), varying in number from three to nine corporate values respectively guiding principles. They usually do not consist of one word (F1; E9) but rather an (imperative) superordinate action title (cf., F2; M6; F10) which is concretized and explained in another sentence. The use of less generic and abstract terms is intended to ensure that, on the one hand, there is less room for interpretation, and, on the other hand, these values can be operationalized in teachable and observable behaviors (F7; F10; M13).

Most founders tend to involve their employees in defining and developing company values and guiding principles, allowing them to view them as "*part of their own intellectual property (...) or their own DNA*" (M6) and accordingly endorse them in their daily actions (F7). This is done through workshops, qualitative interviews, company-wide surveys, or voting.[6] Once the values are written down, some start-ups review them over time to determine whether they are still applicable at the current stage and adjust their wording if necessary. This usually involves only a refinement of the wording regarding the current corporate context rather than formulating a new set of values (F3; F5; M9; F10; M13; F14a).

The corporate values are seen by many founders and managers as the core of the corporate culture, which must be preserved even during growth (F2; F3; M10; M11; M13; F14a; F14b).

Particularly in a start-up's initial phase, however, it does not play a major role in whether a value system has already been written down (F3; F5; M9). Instead, it is essential that the founding team consistently and continuously leads by example from the outset and takes on a (4) role model function, acting accordingly to its espoused values. That is because the employees observe how they communicate, behave and interact with people (F3; F7). Founders and managers referred to this practice as one of the most frequent, significant, and effective one in the cultural context (F1; F2; F3; F5; M6; F7; M8; F10; F12; M13; F14a). Especially in this early stage of a start-up, it is assumed by the interviewees that the founders greatly influence how things are done by (implicitly) exemplifying it themselves to their employees, thus primarily shaping the early corporate culture (F7; E10; F7; F12; F14a) and conveying their basic assumptions (F5; M6; F10; F12). This initial behavior has long-term effects, according to founder F7 and should not be underestimated:

> "What you [as a founder] just exemplify there, the way you behave will be decisive for how the company will behave. If you say, 'Oh, come on, screw it, that one customer there, he's not that important, is he?' or 'Well, we don't have to be so strict about legal', then you always have to be aware that this behavior will shape the company

[6] Exemplary questions start-ups have asked their employees: What kind of culture do we want to have? What makes working at CX special for you? If you like it, why? What makes it stand out? What would be important to you when other people join that they know how work is done at CX? What are the most important values for you? (cf. F3; M9)

> for many years to come because it is your basic attitude. And that's where you have to challenge yourself." (F7)

To act as a credible role model, consistency in all areas of action between what is said and what is done by the founder team is mandatory (M6; F10; F12). Only through this observable consistency, which is evident in their behavior, interactions with other colleagues, and communication internally and externally, do the corporate values attain authenticity (F5; M6; M13; F14b). For instance, one founder stated that if one as a founder has not behaved according to these values, one should make oneself vulnerable and admit the misconduct (F7).

Furthermore, the founders also have a role model function when they point out to their employees the right or wrong behavior in a certain situation which is rather a spontaneous reaction to an occurrence (F2; F3). At the same time, feedback as a management instrument (see Chapter 4.3) is used by all start-ups from an early stage on to regularly address the behavior of managers and employees consciously and more formally and to coach them if necessary.

By creating regular points of contact with the founding team, such as onboardings, training, meetings, and events, or through open office structures, some founders try to continue to provide visibility to their impact as cultural role models despite an increasing number of employees and thus fewer direct close personal interactions (F3; F5; M6; M9). Repeatedly, founders and managers referred to the aspect of the approachability of the founding and management team (F5; M6; M13; F14b).

Several founders reported that from the point at which management levels are established and direct founder contact with all employees is decreasing, it is of great importance to be a role model, especially for employees with management responsibility, so that they can act according to the founders' value system and pass it on to their team (F1; F2; F3; F10) since employees are guided by their supervisor's maxims for action (M10) and observe the behavior of their direct supervisors (F3; M10; M13). In this context, three of the founders emphasize that being a role model is seen as particularly critical for externally recruited managers who may already bring other types of cultural imprints from previous employments with them, as they, in turn, have a direct influence on their team members. For instance, this is not only to ensure that these managers exemplify the start-up's values to their employees but also to prevent undesirable cultural change (F5; M11).

4.2. Personnel Decisions and Onboarding

In addition to the basic entrepreneurial orientation and behavior of the founders, data revealed that human resources processes as well as the onboarding process of new employees are key instruments in the development and maintenance of the respective cultures of the start-ups.

One of the most frequently mentioned instruments for shaping, developing, and maintaining corporate culture was (5) **recruiting**. According to the interviewees, this primarily involves - alongside professional qualification - identifying a person-culture fit on the part of the company and the part of the applicant. Some of the founders and managers attributed exceptionally high importance to applicants' cultural fit, pointing out that applicants meeting the professional requirements but not the cultural ones were not hired (F5; M6; F7).

While in the early phase of the studied start-ups, decisions about new employees tended to be made based on the founder's gut feeling, with the rapidly increasing need for new employees, more structured and multi-stage procedures were introduced in the recruiting process at various start-ups. This is intended to recruit more systematically for a cultural fit.

Already during the initial contact between the applicant and the company, the applicant can get a first impression via the company websites of the start-ups, where usually the company's corporate values can be found or "people stories" (F5), which describe what it is like to work in the respective company. Start-ups C5 and C8, for instance, reported that they had recently revamped their website presence to communicate corporate culture aspects to potential job applicants early on and attract (culturally) appropriate candidates accordingly (F5; M8). The Team Lead Human Resources of start-up C10 emphasized that to maintain their corporate culture in hyper-growth, it is even essential that applicants are familiar with the start-up's values from the beginning, as this would allow them to decide if this is a culture they want to work in (M10).

When asked how the different start-ups determine a cultural fit, respondents mostly replied that involving multiple people in the recruiting process and getting to know each other can develop a good sense to evaluate whether an applicant is a cultural fit for the start-up. For this reason, the application process is typically multi-step for all start-ups and involves various employees to assess the applicants' fit with the corporate culture. For instance, several peer sessions and team interviews (M6; F14b), and a decision panel (M11) that decides whether to hire an applicant were mentioned. Moreover, C13 utilizes reference calls to confirm statements made by the applicant (M13). C3 has introduced an all-day experience day early on as a last recruiting process for full-time employees. The purpose of this day is for the applicant to experience the start-up's culture and for the team to find out whether there is a team and culture fit between them and the applicant. Thus, the team makes the final decision on whether to hire an applicant. F3 explained that it is feasible that the previous interviews with HR and the team lead were positive for the applicant, but the team decided against the candidate (F3).

In the interviews themselves, situational questions, for example, are aimed at verfiying verify whether the candidate has acted and operated according to the start-up's values in previous employment and whether they support the start-up's corporate values (M4; F5; M9; M11).

There was agreement among the interviewees that corporate culture is a significant and fixed evaluation criterion in recruiting and is thus included in the candidate evaluation form or scorecard, for instance (F2; F7; M9; M11).

In all start-ups, up to a certain employee size, at least one of the founders was involved in the final recruiting interview, which is also seen as a kind of "*values interview*" (F10) or "*culture-fit interview*" (F5). At some point, however, some of the founders have withdrawn entirely from the recruiting process due to other obligations. In this regard, one of those founders reported the following:

> "In fact, until recently, it was the norm that one of the founders must have interviewed every employee. In the meantime, I don't interview everybody anymore, [but the hiring manager does]. (...) but there is much risk in that. At some point, I will realize that I should have done this for a longer time. But it's always a question of prioritization." (F7)

In fact, some of the founders of the sample attributed a particularly critical role to the staffing of HR positions, which are intended to assess cultural fit on behalf of the founders (F3; F7; F12). Nevertheless, even while growing rapidly, founders continue to be part of the final interview at nearly half of the start-ups (F1; F2; F3; M4; F5; F7; M13; F10), although in some cases only for management positions (F3; M4). As long as capacities allow the founders of C2, C10, and C13 aim to remain involved in the recruitment process of their employees while scaling.

When (6) **selecting** employees, a specific "*type of employee*" (F5) is sought, with the right "*mindset*" (M6; F7; M8), similar cultural perceptions (F1; F2; F7; M8), and which "*tick*" (F2) similarly to the founders. Accordingly, it could be observed in some of the interviews that, especially at the beginning of the ventures, friends and former colleagues were hired in which a cultural fit was seen by the founders, who in turn referred other employees (F1; F5; M6; M11). The Chief of Staff of C6 and COO of C11 attributed a significant contribution to how the corporate culture has evolved to these individuals through their shared history from a previous employment relationship. The two founders of start-up C15, on the other side, emphasized that it was crucial to fill the first position to be assigned with a People and Culture expert to have the appropriate expertise in-house from the outset to develop the type of corporate culture they were aiming for.

The interview data revealed that it is not a matter of selecting people who resemble the founders in their characters or traits but rather bring a similar set of principles and attitudes, and share the mission, thereby creating diversity in the workforce nonetheless (F2; F3; M8; F14b).

The responses of the founders and managers further indicated that, particularly regarding the rapid growth of the start-ups and the associated changes, the characteristic of willingness to change in a corporate context is sought in the employees:

> "I think that the topic of hyper-growth requires a very special mindset. We have only hired people who are insanely willing and enthusiastic about change." (M8)

Respondents explained that this is because start-ups are subject to constant change, which means that processes and structures are constantly changing internally, and adaptability is required accordingly (F1; F2; F3; M8; M9; M13).

Being very restrictive in selecting employees from the very beginning was seen as a prerequisite for shaping the desired type of corporate culture (F2; M4; F5; F7; M8) and fundamental to maintaining it (F2; M4; F5; M6; F7; M13). One interviewee pointed out that hiring the wrong person could "*quickly destroy the corporate culture*" (M4). A very restrictive selection approach must be maintained, even if the rapid growth is dominated by a high demand for new employees, to preserve the corporate culture at its core:

> "Being very restrictive in the selection of employees is why we have managed to maintain the corporate culture." (F2)

Almost half of the founders and managers analyzed in this study reported their (7) **onboarding process** as a particularly powerful instrument in preserving the start-up's culture as it grows (F1; M6; M9; F10; M13; F14a). They described addressing their corporate culture there in various ways. In general, the onboarding is intended to help integrate new employees into the start-up's culture more quickly.

Some of the founders and managers emphasized that as the number of employees grew, parts of the onboarding process had to be adapted and expanded and new formats introduced to ensure the successful onboarding of new employees during such rapid growth, also considering the remote-work context imposed by Covid-19 (M6; M9; M13).

Blueprints for different onboarding journeys, onboarding sessions, and onboarding classes were mentioned, for example, where new joiners are taught cultural elements [7], and the introduction of buddy programs with at least one buddy per new joiner (M6; M9; F10; M13; F14a, F14b). In a buddy program, an experienced employee is assigned to help a new hire become socialized. The buddy shares his or her knowledge about the company, and specific questions about the culture can also be directed to the buddy. The buddy thus also serves to communicate the corporate culture explicitly and implicitly (Graybill et al., 2013, p. 203). The CPO of the fast-growing start-up C13 sees the advantage in onboarding classes that the newcomers gain a cross-departmental perspective, have direct points of contact, and possibly grow together as a group. Company C9, which already has several offices globally, initially holds a global "*Culture Onboarding*" to convey the idea "*We are one company, and we have one corporate culture*" (M9) to the new starters from the various

[7] For instance: How do we work? What is important to us? How do we deal with each other? (M10; M13)

offices. This is followed by an "*Office Onboarding*" that addresses local peculiarities.

Among start-ups, the onboarding process ranges from a few days, weeks to six months[8], whereby some study participants emphasize the importance of in-person onboarding to help newcomers get a faster impression of the nature of the start-up's culture (F2; M13; F14b). For example, the co-founder of start-up C2 explained that even if new employees start working remotely, the first two weeks are mandatory for them in the office, and again after six months (F2).

Parts of the onboarding process deal with the topic of corporate culture in a dedicated way, e.g., in the form of a culture day, culture sessions, and a core value training in which the issue of corporate culture is explicitly addressed and discussed.

A key component of onboarding that has usually been part of it since the early days of each start-up is the *founder touchpoint*, at least in one part of the onboarding process. The founders welcome the new joiners, tell the company history and founding story, present the corporate values and guiding principles, and explain what these mean specifically in the company context, how and why they were selected, and what is expected in this regard (F2; F3 M9; F10; M11; M13; M12). In some start-ups, such as C2, C7, and C9, care is also taken to establish a personal connection to the values, e.g., by asking which values the new employees identify with most and why or simply creating a platform for discourse about the corporate culture and values (F2; F7; M9). From the founders' perspective, the onboarding process significantly contributes to conveying their basic assumptions (M4; M8; F10; M11; M13; F14a).

According to M13, it is essential to keep the founders as a central figure of the start-up's culture involved in the onboarding process, even if the number of employees increases significantly. Correspondingly, the format of getting to know the founders of C14 has been adapted as the rapidly growing number of employees was no longer manageable with the previous format.

Instead of half an hour for each employee to get to know each founder, there is now a one-hour call with all newcomers per month with all members of the founding team.

The CPO of C13 reported, that she is currently working with one of the founders to redesign the onboarding process due to their rapid growth and reported the following:

> "We want to double again next year. You can't grow that fast without having a very strongly structured onboarding process, where exactly all these cultural elements are communicated uniformly and consistently: what do we stand for, what are our processes? (...) And of course also so that employees can now be onboarded remotely. That was a huge challenge for all companies in 2020, because it's quite normal to say, 'Hey, we're going out for a beer, here's a coffee'. How do you manage that? I.e., you also have to go much more for such a visual language. You have to start doing video content. That means we are building a tool-supported, structured onboarding process to penetrate this (cultural) theme." (M13)

This example summarized once again that as the number of employees increases, the onboarding process is adapted, and the need is seen to explicitly address the issue of corporate culture there.

Another measure to develop and foster corporate culture mentioned frequently by the respondents is the (8) **promotion** of early joiners, i.e., the first employees of a start-up (F3; F5; M6; F10; M11). One reason they referred to is not to lose the first employees in the long term since they are usually strong culture carriers and anchors in the eyes of the founders:

> "We have always been fortunate enough to have had quite strong [cultural] anchors by splitting the early joiners into different leadership positions in different teams." (M6)

Having early joiners in higher positions "*leads to a better communication of the culture because they have known the culture for longer. They will also tend to lead their team in this way.*" (F3)

The COO of C11 expressed the importance of the promotion as follows:

> "Quite a few of the team leads and Head ofs that we have in our company are employees who have been with us for a long time, have absorbed [the culture], were successful in their way, both in results and values, were [promoted] based on that. They then have a role model and spreading function throughout the organization." (M11)

With the introduction of a management level, emphasis was placed on communicating to those managers that they also have a supporting role in transmitting the start-up's culture. Employees who have been developed into managers thus also hold the responsibility of role models and act as multipliers of the corporate culture (F3; M6; M11). Furthermore, the COO of C11 stated in a previous interview that in his view promotions "*are [in general] one of the strongest culture-building signals for the organization about desired behaviors and results*" (Martinetti 2020, External Interview with M11). In addition, internal job postings are published internally first, giving existing employees the opportunity to apply for an open position before external applicants can do so (M11).

[8] The six-month onboarding process involves weekly sessions that new employees must attend a certain number of times. There, people deal with and discuss various aspects of the corporate culture. At the same time, the newcomer learns what is essential to the company in terms of working methods and mutual interaction. (M8)

This is countered by the (9) **sanctioning and separation** of employees who do not behave in accordance with the type of corporate culture the founders aspire to or who violate corporate values (F2, M13; F10; F14a; F14b).

Another reason that was observed from the interviews that can lead to the separation of employees in the context of corporate culture is when the company outgrows an employee (F5; M13; F14b), i.e., when the entrepreneurial changes associated with rapid growth of a start-up are not embraced and supported by an employee.

4.3. Evaluation of Employee Behavior and Feedback

The evaluation of employee behavior in the form of appreciation, performance assessment, and feedback were also applied as instruments for culture development and maintenance.

Linking (10) **awards and appreciation** to the company's corporate values and the behavior founders are concerned with is another instrument that is used in many start-ups.

What was initially done informally, directly person-to-person, occurs at some start-ups in a more formal, systematic way after a few years and with a larger number of employees. Again, the start-ups' execution varies. For instance, two of the founders report how employees are praised in communication channels for their behavior in relation to the start-up's values, both by founders and other employees (F5; F7). At C5, there is even a designated "*appreciation channel*" on Slack, the internal communication tool. The idea is to praise the person who has exemplified a corporate value particularly well and thereby express appreciation for the behavior. At the same time, other colleagues are made aware of which behavior is highly valued in the company and how one has to behave to receive (public) recognition for it. In the same way, it is pointed out when a value has not been acted upon or should be observed in a given situation (e.g., the value "focus") (F7). Other founders in the sample also emphasize the exemplary behavior of their employees in meetings:

> "What we focus on very strongly is simply showing recognition and appreciation for people (...) by actively highlighting people in larger meetings with reference to these values, where they have acted accordingly in the specific situation. I think that is very, very important for everyone individually." (F2)

In addition, four founders mentioned awards as a more formal type of recognition, which are handed over to employees on a (bi-)weekly or quarterly basis (F2; F10; F11; F12; F14a). This usually follows some sort of process that requires the employee to be nominated for the award. However, employees do not need to be nominated by the founders themselves or their respective supervisors; other employees can also nominate them. At C11, for example, a jury decides who ultimately receives this award, whereas at C14, the employees vote for the winner via an online voting tool.

Some founders and managers highlighted, when publicly awarding employees, the audience is told why this person won and how they behaved in a particular situation. This not only expresses appreciation to the winner but also serves as a "*signaling effect of what is desired here in the organization and what great effects it can have if you behave that way.*" (M11). Sharing the personal stories around each corporate value or principle conveys the meaning of those into the company. Start-up C15 also publishes in a LinkedIn post an example of an employee who received an award in the current week, which value he or she exemplified particularly well, and why[9].

Numerous founders and managers also referred to (11) **performance assessment** as a very explicit instrument for maintaining their corporate culture as performance can be linked to the start-up's values there.

In semi-annual or annual performance reviews, in addition to the performance achieved, there is a review (of equal value) of how employees have lived and "*delivered*" (M13) on the corporate values "*so that everyone's behavior is repeatedly reflected against this canon of values*" (M13). Founders and managers outlined that specific situations are inquired in which individuals have behaved in accordance with the values in the work context; they are evaluated on the basis of these (M4; F7; M11; M13). As a result, "*alignment with the values is also linked to compensation and career progress*" (M11), and each employee is "*accountable*" (M13) in this regard. The extent to which employees pay attention to the corporate values is assessed at C13 from interns to high-level management (C-level). The founder F14a named the American human resource software BambooHR as a specific software tool for performance evaluation. On its website, the technology company BambooHR explicitly addresses the topic of performance management under the tab "culture" and states that performance management directly influences corporate culture (BambooHR, 2022).

Besides the performance assessment of employees, all start-ups from the selected sample build upon a distinct feedback culture. Systematic feedback collection on the company is discussed in Chapter 4.10 Active Participation and Involvement of Employees. This section deals with frequent bi-directional (12) **feedback** between employees of the same level and between employees and their superiors. The focus of this instrument is not on the assessment of an individual but on a regular exchange, which is intended to help promote desired behavior within the company, personnel development, and capturing employee concerns and worries and general suggestions for improvement. Accordingly, different methods of conducting feedback were mentioned: (annual) upward manager feedback (M11; M13), where supervisors receive feedback collected from their team members, peer feedback (F3; M11) between team members, (weekly, quarterly) one-on-one feedback between supervisors and their employees (F2; F10; M11), and retrospectives (F10). Likewise, feedback occurs in a more informal setting by directly

[9] An example can be taken from Appendix D.

pointing out situationally when an employee's behavior does not meet (cultural) expectations (F1; F2; F14). The different feedback methods communicate which behaviors are (not) desired[10], whether employees act according to the corporate values, and what kind of corporate culture is aspired by the founders and managers (F2; F7; M11):

> "We constantly have feedback, which practically also represents how we see ourselves and how we see the culture, that we have high standards for ourselves, but that we also want to help others to become better. And that's why we give and take feedback." (M11)

4.4. Resources

This section explains resources that are used in the development and maintenance of the company culture (training, workshops, and time and human resources) or reflect cultural values in the use of resources (salary and benefits).

Financial compensation, whether (13) salary or benefits, was cited as another example that is linked to company values (M11; F12; M13; F14a). At C12, the bonus system of the sales team reflects which behavior is desired:

> "(...) In sales teams, where we naturally also have a certain competitive situation, and also honestly promote it, we also attach importance to (...) how our bonus system is designed. (...) For instance, if one of [the sales team members] is sick or on vacation, that people help each other out and support each other and don't go for the elbow mentality." (F12)

If it is essential to one of the start-ups that all employees with the same job are treated equally, then the salaries are dynamically adjusted accordingly if the labor market demands increasing salaries and new employees receive a higher starting salary for this reason than employees who have been with the company for a more extended period (M11). The assumption of equality may also manifest in the non-existence of salary differentials between genders, as in C8. Therefore, in essence, compensation, on the one hand, is instrumentalized to influence the behavior of employees and, on the other hand, expresses the values and principles of the start-up.

Another instrument to convey what is important to the founding team and what kind of behavior is desired is (conceptual) (14) **training and workshops**. This more formal type of manifestation would be required in the case of rapid and continuous personnel growth and in the assumptions of doubling or even tripling within a year (M9).

Through training, learning development programs, and the associated financial and time investment, the start-up expresses what it considers important in terms of its corporate culture (M8). Examples given are "*empathy communication*", "*agile working*", "*remote working*", "*team effectiveness*" or "*self-leadership*" training (M8; M9). Accordingly, C9 has created an in-house training and learning offering designed by the internal learning and development team "*to develop the mindset and competencies needed to drive system-change for sustainability effectively*" (M9). Furthermore, to counteract a possible cultural decline due to rapid growth, a specific training was developed with more formal reference to their corporate culture (M9).

Especially importance is attributed to training for the management level (M6; M9; M11; F14b). Such training is supposed to be a particularly scalable and effective instrument since managers have a greater cultural impact merely by their position:

> "By far the most scalable thing you can do (...): you have to make sure that the managers in individual teams [are well trained], know [what is important to us founders], how to communicate these values, what we expect from them, what kind of role model they should be, (...) to align them with our values." (F14b)

In addition, training courses and workshops are also offered and conducted that directly address the corporate culture itself, some of which are already anchored in the onboarding process of many start-ups. These include training on the start-up's corporate values - "*What do they mean? How were they selected? What is expected of employees in this regard?*" (M11). Usually, training courses that do not directly deal with the company values are not conducted by the founders themselves. But according to M9, this would have the advantage that participants could learn directly from the founders and ask them more in-depth questions about the content. Such a format involving the founders could directly influence the "*next generation*" of new employees (M9).

4.5. Organizational Design and Structure

Only a few start-ups initially associated (15) **organizational structure** and how the organization and teams are built as an influential instrument for shaping corporate culture (M9; F12; F14a; F14b). However, once the term was introduced into the conversation by the interviewer, this instrument was given substantial weight to influence the behavior and mindset of employees by the founders' assumptions and values about how to conduct things as the organizational structure reflects the way internal relationships and communication is managed, roles and power are allocated (F2; M6; F7; M8; F14a). For instance, at C9, a consultancy, a distinction is only made between associates and partners, which keeps the hierarchy level as low as possible. At C8, care was taken in the organizational structure to separate technical and disciplinary leadership, indicating how leadership is conceived within the start-up. Many start-ups studied hold the fundamental assumption that core competence resides in their employees, which is why decision-making power is

[10] "If you somehow have the feeling that people are not acting [in accordance with the company's values], you don't just let it go, but address it in our weekly one-on-one meetings, which always take place between employees and their supervisors." (F2)

transferred as deeply as possible into the organization and responsibility is delegated (F1; F2; F5; F7; F14a). To maintain specific values in growth, such as the value of "Ownership", the organizational structure at C15 was adapted accordingly.

One section of the interview dealt with whether there is a (16) **dedicated function** internally for the topic of corporate culture at the company size currently obtained. It was observed that there tends to be a consensus that the founding team bears the primary responsibility for the start-ups' cultural topics, as it is "*possibly the most important job [of a founder]*" (F2; F10).

Nevertheless, many of these founders described that the shaping and implementation of development and maintenance activities are primarily assigned or located to a department such as People and Culture (F7; M10; M11; F12; M13; F14a). In other start-ups, there are institutions such as a global culture committee (M9), a culture club that every employee can join (M6), or a normative circle, which includes, for example, employees from the Human Resources and Legal departments, Agile Coaches and people elected by employees (M8), where corporate culture topics are talked about in a dedicated way. The founders may be represented in these institutional groupings, though (M8; M9).

The importance of a group of people dedicated to the topic of corporate culture from a particular company size - because it is a too "*important a topic*" and should not be a "*side topic*" in the eyes of many founders (F5; F7; F14a) - is also shown by the following quote:

> "[Anne from the People Team] unfortunately left the company a few months ago (...). That's why we have to anchor [the topic of corporate culture] more strongly again. Because if you have a person like her who also owns it, you also have an effect behind it. (...) But [aspects of corporate culture] will bleed if we don't find someone who really takes it up again and sees it as a central task." (F7)

The first employee at C15 was not only assigned to their People and Culture department, but the start-up generally followed the approach of "*really (...) investing in the People and Culture department*" (F14a) and assigning one HR Business Partner for every five to six managers since the management level was introduced. According to F14b, they assisted as a strategist in guiding the company through the various growth phases from ten to 200 employees and made this rapid growth within 18 months possible for the start-up in the first place. The founders have not encountered this staffing ratio in any start-up before (F14a; F14b).

Besides C1, the two hyper-growth start-ups C4 and C5, also plan to hire a person such as a Head of People and Culture (M4) full time due to the very rapid growth in personnel, who will be "*hired specifically to [further] build [C4's] corporate culture*" (M4) or "*deal with the issue [more strategically]*"(F5).

4.6. Organizational Processes

In general, the internal (17) **process design** as a cultural instrument was also given relevance that permanently recalls and positively reinforces the corporate culture (M11). The interviewees refer to specific management processes such as recruiting, onboarding, performance assessment, or feedback processes, which have already been discussed in previous chapters, or more general statements are made about management processes:

> "In addition to these very obvious reinforcers of values, we also try to design internal management processes in such a way that we have the feeling that this is a coherent image [with our culture]. (...) And then to ensure that [the corporate culture] is always remembered and positively reinforced through practically sustainable, scalable processes, whether that's through the annual reviews or value awards, but the most important thing, I think, is to pay attention to it in recruiting." (M11)

With regard to growth, not only structures but also processes must be constantly reviewed and rethought in terms of their suitability (M6; M9).

4.7. Events and Rituals

The reflections of the founders and managers showed that rituals, i.e., recurring events and activities, are particularly important for the development and maintenance of the start-up's culture (F2; M6; M9; F10; F12; F14b). Ritualized activities are not unique to this chapter; for example, giving awards, which have already been addressed in a previous chapter, can also be considered ritualized activity. The focus of this chapter is on creating further regular points of contact between employees, where the corporate culture should not only be experienced by the individual but, in some instances, also explicitly talked about.

Various founders and managers attributed particular relevance in a cultural context to internally held (18) *events* that strengthen *team cohesion* and encourage exchange between employees on a regular basis (F5; M9). This is intended not only to counteract the anonymity that threatens with an increasing number of employees (M4; M11; F12) but also to reinforce the corporate culture by embracing and experiencing it there in a more informal setting (F2; M6; M10; F12; F14a). Creating emotional experiences through events promotes cohesion among employees and fosters behavioral patterns (F10; F14b). Besides, new shared company stories (M6; F10) are created that transmit the start-up's values (F10).

Joint activities and teambuilding measures take place within or outside working hours. Especially in the initial phase of a start-up, such activities increasingly take place spontaneously and informally (F5; M6), i.e., in the form of spontaneous get-togethers after work in the office or outside

of it. Such activities as spending the evening together or going out to celebrate should not be underestimated, as "*they also imply cultural elements*" (M6).

As the number of employees increased, the various start-ups placed emphasis on initiating even more regular networking opportunities, both within teams, across teams, and as an entire company (F1; F2; M4; F5; F10; M11; F12; M13). Events outside of working hours tend to be based on voluntary participation, are less formal, and have no agenda. Examples that were mentioned among several other networking opportunities currently used by start-ups included (online) randomized coffee, breakfast or lunch dates (C4; C9; C10; C11; C12), speed dating (7 dates in 7 minutes) (C2), movie nights (C11), and joint sports activities (C11; C12). This also includes company parties such as Christmas parties, summer parties, etc. (C1-C14).

It is apparent from the interviews that during the Covid-19 pandemic and the accompanying contact restrictions, on-site events could only be replaced by remote events to a limited extent (F3; F5; M6). Nevertheless, it was important for the start-ups to develop online alternatives or other in-person formats, allowing smaller team sizes to gather. One digital format introduced during the pandemic and which has proven successful for some start-ups is the randomized coffee dates that take place online between (at least two) employees from different departments or locations (M4; M9; M10; M11).

A few start-ups also mentioned (19) **events** that are directly **dedicated to corporate culture** or events that explicitly address the start-up's culture in sections, although these would only become relevant once the company reaches a certain size (F14b). In three cases, reference was made to the "All Company Culture Week" (#ACCW) of the German Unicorn start-up Personio, which already had over 750 employees at the time of this event. In summer 2021, the start-up invited all employees from across Europe to its Munich office "*to reaffirm [their] culture and grow together. #ACCW was designed as an opportunity for all Personios to come together, meet in person, and be fully immersed in what [they are] about and why [they] work the way [they] work.*" (Personio, 2021). In addition to many activities that made the corporate culture tangible[11], the start-up's culture was also explicitly discussed in a culture-focused Q& A session and the entire "Values Day".

So far, only a minority of the start-ups from the sample have conducted a similar event on a smaller scale. The founders of C15 are currently planning to organize this type of event for its nearly 250 employees located in several countries as soon as the Covid-19 situation allows (F15b). C11 is also planning an event under the slogan "Culture Days" for its more than 330 employees. Besides, at C13, the entire company with its 180 employees is invited to a different location once a year for one week to "*experience the cultural spirit*" (M13). Another founder considers the idea of such a culture week, as conducted by Personio, to be basically beneficial. However, the associated financial costs to bring all employees from the different locations together are too high in his opinion (F3).

Since its founding, C9 has held the C9-Day twice a year, a one to two-day strategy offsite to which all employees are invited to a single location to talk about the company's mission, impact, projects, and culture.[12] However, this event is not solely for the purpose of addressing corporate culture. Instead, it is a platform that offers the opportunity to explicitly address corporate culture and bring all employees together in a ritualized way. This had already been done in the initial phase, which is why this format is seen as particularly relevant for developing the foundation for the desired corporate culture, according to M9, but also to later introduced cultural initiatives for maintaining it.

In addition, the analysis of the collected data revealed that recurring (20) **meetings** are an instrument for founders to express corporate culture explicitly in some instances or to make it more tangible by behavior and thus reinforce it. What initially required less explicit expression, as long as "*all employees still fit in one room*" (F2; F3; M8; F7), is seen as necessary as the number of employees increases. Company C2 shows a very straightforward way of addressing the topic explicitly in meetings at present:

> "We always have our team weekly, once on Friday, where the whole company gets together, and we go through the updates of all departments. There it's important to demonstrate the cultural matters again by highlighting people (...). We always discuss or talk about the company values for 10 to 15 minutes at the beginning of the meeting and then ask all the people to actively bring examples of this, so that we are not always the sole entertainers, but that the employees also bring examples of this. This also gets them used to somehow highlighting other people in front of the large group and addressing these values." (F2)

This very detailed example shows that the corporate culture or formalized elements of it, as the start-up's values, are, on the one hand, formally on the meeting agenda and, on the other hand, visibly exemplified for all employees. Other start-ups also try to link the company values with the topics of the meetings (F2; F12; M13), e.g., to illustrate the extent to which one of the topics being worked on at the moment pays off in terms of one of the company values (E13).

Praising exemplary behavior and allocation of value awards, which have already been discussed in more detail in Chapter 5.3, usually take place in the regular "*all-hands*" (M4; M11; F12; F14a) or "*town-hall*" meetings (F2; M9).

[11] One of the company's values is social responsibility, which is why every employee has two paid days per year to get involved in a cause that is important to the employee. During one day of #ACCW, employees were able to work on an Impact Project.

[12] Since Covid-19, this event took place online.

This meeting format brings together all the employees and management of a company to present relevant updates and information, clarify related questions[13], and address goals and corporate values. One of the start-ups with over 300 employees spread across different locations holds additional local town-hall meetings to share more specific information that pertains only to that location (M9).

In most of the start-ups in the sample, these types of meetings are held primarily remotely, weekly or bi-weekly, or monthly in the case of C9, with more than 300 employees. M13 believes that all-hands meetings should be held at least every two weeks. In another start-up where she previously worked, all-hands were held monthly, which she felt was insufficient.

Topics and the way they are presented in those meetings also reflect the founders' values. For example, "*transparency*" plays an essential role for many founders (F2; F5; M10; M11; M13; F14b), which is why they report on the "*bad and good company news*" (M11) in such meetings:

> "We create value for our customers by bringing transparency into the [..] market, and we also believe that internal transparency builds culture and business. For example, in our bi-weekly all-hands meetings, we transparently share the most relevant business figures and give timely updates on good news and bad news. We believe that if everyone is aware of the objectives and challenges, we can source the creativity and knowledge of the entire company to get ahead." (Martinetti 2020, External Interview with M11)

The way meetings are generally held and are structured also reveals parts of the corporate culture. For example, one corporate value of start-up C3 is "*use humor*"; according to F3, it is "*completely okay if jokes are made even in serious meetings, which is not planned, but simply arises situationally*". On the other hand, in the case of C6, observers would quickly realize that the interaction between employees in meetings is very casual but also structured. However, these are not "*fun meetings*" (M6).

Other ritualized activities emerged from the interviews, such as sending postcards (M9; M10) or welcome and care packages (F1; F3; F10; M13) to employees. Since these rituals were not covered to the same extent in the interviews as the previously mentioned aspects, these points are mentioned for the sake of completeness but are not assigned equal relevance.

4.8. Office

Whether their architecture, spatial design, or utilization, the offices were also regarded as a factor influencing corporate culture in the work environment by the founders and managers.

The fact that deliberate attention is paid to the (21) **office layout and design** so that the corporate values of the start-ups are made more visible as a result emerged only from a few interviews (F3; F5; M6). The aim is usually not to only put the corporate values on the walls (M6; F12) but to make them tangible or observable through the office layout and design. A significant corporate value at C3 is prototyping, which implies that "*things are built*" in their corporate context, and production is essential for the core business. For this reason, an office was deliberately chosen that does not separate production from the other departments but locates it on the same level as, for example, the engineering and sales departments. In addition, there is a large workshop so that tangible prototypes can be built directly during the generation of ideas if required. In general, different prototypes and hardware can be found throughout the entire office (F3).

If proximity to their employees is central to the founders, this was reflected by the fact that the founders do not have individual rooms but rather locate themselves together with their employees in the open workspace:

> "People have already come into the office and asked us [founders] why we don't have our own individual rooms and sit in the open space. But that's a no-brainer for us, that we sit together with our employees, because that allows us to exchange a lot and because we also want to be close to the team and stay close." (F5)

A similar approach could be observed at C6. Such reflections indicated that the physical context also strongly influences the flow of information and communication and the interaction between employees.

According to two interviewees, the office should be designed in such a way that people feel like spending time there, which also promotes exchange again (F5; M13).

In addition to the office's design and layout, the (22) **office utilization** was also described as a possible instrument to strengthen the development of the corporate culture. Even though many of the start-ups are now pursuing a "remote-first"[14] mindset, which "emphasizes equality of access and ensures that there are no advantages or disadvantages to working remotely" (Gold, 2022)[15], the majority of the interviewees emphasized that the office is still a central anchor

[13] C10-C13 use the "AMA" or "ask me anything" method in this context, which enables their employees to ask questions live during a meeting anonymously, which is necessary above a certain company size (F10; M11; F12; M13). In doing so, meeting participants can vote for questions relevant to them. The questions with the most votes are then answered directly live by the founders or management. Slido, a Q&A- and voting platform for meetings and events, was mentioned as a suitable tool for this method (M11; F12).

[14] Other designations: "remote-flex-policy" (M13), "digital-first" (F2). This development was mainly driven by Covid-19. However, this mindset does not imply remote only (F14b).

[15] *"No matter where people are working from, they always have the same opportunity to participate in a meeting. (...) [That means] even if you are dialed in remotely, and several people are in the same room, everyone dials in with their own camera so that the person who is not in the office does not have the feeling of being the fifth wheel but can participate in the conversation just like everyone else." (F2).*

point for gatherings, workshops, and management meetings. However, these meetings currently take place rather selectively and prospectively due to the continuing pandemic. On the other hand, other start-ups strive to remain an office-led company where the workplace serves in the work context and outside of actual working hours (F3; F5; F10; M13). The intention behind this is that the company members also have the possibility to exchange ideas in more informal settings and that corporate cohesion is strengthened (F2; F5; M6; M13). The example of F5 also illustrates this:

> "We also have much space for leisure activities; there is a PlayStation, a couch area, table tennis, darts, and a pretty cool kitchen. And I think that's also super important for the team culture. Of course, that's more the leisure part and not the work part. But I think it's important to have a workplace where people like to come and spend time." (F5)

From the perspective of the founders F3 and F5, the corporate culture is difficult to grasp for people who are not in the office.

If several office locations already exist, founders and managers from the headquarters (23) **visit** them to convey to them "*how the company works*" and the desired corporate culture, but also to get an impression of how the location there is developing culturally (F2; F3; M4). Such visits were considered as a promotional practice, especially in the establishment phase of new offices (F2; M4). For this reason, members of the founding team at C2 regularly spend time in other offices. In the same way, employees of those offices are invited to the headquarters to gain an impression of how the working methods and culture are there. Despite the visits, there is no insistence on imposing the corporate culture of the headquarters one-to-one on the other offices.

Some freedom in cultural development is allowed in different locations as long as the core value set is the basis for it (F3; M9; M6).

4.9. Communication

In addition, the interview data revealed that consistent and continuous communication about the start-up's corporate culture and the way it is communicated have an influence on the development and maintenance of their corporate culture during growth. Three key aspects emerged from the interviews: Writing down and documenting corporate cultural mindsets, the communication style, and communication channels used by the respective start-ups.

One pillar of corporate culture communication is the (24) **written formulation and documentation** of cultural mindsets. In this context, in addition to the vision and mission statements, the corporate values in particular are used for communication in the start-ups to establish them as behavioral guidelines for the members of the company.

As mentioned in Chapter 4.1, all but one of the start-ups have written down values and/or the guiding principles derived from them. During the founding phase or the hiring of the first employees within the first year, some start-ups had already formulated these (F2; F3; F10; F14a). In contrast, some other start-ups did not see the need to write down their corporate values or principles at the beginning because "*so much of it is taken for granted*" (E9), and they were already lived without putting them into explicit words (F3; M8). However, codification of those values and principles was usually undertaken once the workforce numbered 30 to 80 employees (F1; M4; F5; M6; M11; F12).

> "I think that from a size of 30 or more employees, corporate culture has to be formalized. Before that, [founders and the very first team members] have a very strong informal work, and everyone kind of perceives the culture that way. But I believe that from 30 or 40 employees at the latest, it no longer works if you don't formalize the culture and consciously implement it. Otherwise, I believe there will simply be a proliferation because this informal communication through a role model function, through perception, functions less and less because the touchpoints are becoming much fewer. And the larger the organization is, the more important these formal topics probably become to ensure scalability." (M11)

All founders and managers agreed that writing down the corporate values and hanging them on the office's walls is of little use if they are not lived (F3; M4; M8; F10; F14b). Albeit documenting and articulating these core values makes them visible to both management and employees, allowing them to be claimed for and used in discussions (M6; F10; F14a).

It was apparent from the data that acceptance of the value statements is strengthened when employees are engaged in the process of defining the core values. Involving employees can also ensure that these values adequately reflect how the company actually operates (F3; F7). Whether the entire company is consulted or (team) representatives are involved in the formulation process, varies (F1; M6; F11; M13; F14a). In the formulations, care was taken to ensure that the company's canon of values is translated into concrete guiding principles or maxims for action rather than generally applicable buzzwords that leave too much room for interpretation (F7; M8; F10).

Once the values have been written down for the first time, they are reviewed at certain intervals to confirm that they are up to date and still suitable for the rapidly developing company and, if necessary, adapted to the current circumstances of the start-up (F3; M13; F14b). It is somewhat unusual for the corporate values and guidelines to deviate particularly strongly from the formulated initial versions. Instead, adjustments and refinements are usually made to the formulation over time.

In addition to corporate values and guidelines, documented cultural artifacts can also be found in articles in the founding convention, employee contracts, code of conducts, and policies (M4; M6; M8; M10; F14a). Managers at start-ups C4, C8, and C10 believed that implementing a code of

conduct and policies is necessary as the workforce grows rapidly and multiple locations are opened. The involvement of employees in developing a formulated code of conduct could be observed here as well. It conveys the culture, values, and general understanding of how work is done (M10).

The documentation for articulating corporate culture can be found at start-ups in employee handbooks, wikis, or intranets (F2; M6; F7; M9; M11; M12; M13).

The interviews repeatedly pointed out that the way of communicating, and thus the (25) *communication style*, contributed to the start-ups' ability to maintain a strong and positive corporate culture during rapid growth. There was frequent reference to *transparent communication* in this context (F3; F5; M10; F12; F14b). This concerns not only explanations of the cultural mindset of start-ups, but also company updates (*"Where does the company stand, what successes and setbacks do we have? What is [the management team] currently thinking about?*" (M13)).

To reduce uncertainty among employees during rapid growth or in the event of an upcoming major change for a start-up such as an IPO, as in the case of C8, the founders communicate transparently and continuously about these forthcoming changes and the reasons behind them in the sense of expectation management (F3; F5; F12). In a fast-changing environment in which start-ups find themselves, it is not only significant to provide information about upcoming changes and the reasons behind them, but also to engage in dialogue with their employees through various formats, to be approachable, to listen to them, and to capture any concerns and expectations (F3; F5; M6; M9; F11).

> "We planned the division of the team a very long time ago and communicated to the team that we were planning to do this, obtained feedback well before we did it, and discussed potential problems or concerns with the team. That is actually the most important thing, that you get everyone involved and make transparent what you are planning and what growth means for the team in particular." (F5)

Next, there are the (26) **communication channels** used in the start-ups that shape their corporate culture and through which, from the respondents' point of view, the corporate culture is promoted in terms of maintenance. The communication channels have changed considerably over the growth phase as they no longer suited the size of the company (M3; F7; M8; M11; F12). The following focuses on the status quo.

Almost all respondents cited *face-to-face communication* as an effective communication channel to convey the start-ups' corporate values implicitly or explicitly, including meetings, events, informal gatherings (see Chapter 4.7), and feedback (see Chapter 4.3). Founder F10 and Chief of Staff M6 explained that, especially in an informal setting, there is an opportunity to listen to employees, seek feedback and address their concerns when needed. This communication channel has been affected by the Covid-19 pandemic and, in the eyes of some interviewees, cannot be replaced by remote formats, or only to a limited extent (F1; F3; M6). Others, however, do not see remote communication caused by the pandemic as too much of an obstacle to their corporate culture. They have taken a digital approach from the earliest days and thus prior to the pandemic (F5; F14a).

All start-ups also use *instant messengers* as another internal communication channel. For instance, most start-ups use Slack, a web-based instant messaging service, for internal communication. It is also used to express corporate values in dedicated channels such as the "*Thank You*" or "*Appreciation*" channel (F5; M8). Another finding was that the more employees the start-up has, the more there is a need to communicate cultural content in visual language (M11; M13), i.e., through photos, videos, and for instance, GIFs, especially in increased remote context:

> "What has been extremely helpful in the meantime are GIFs [that we created for each of our guiding principles]. Whenever someone has acted in the spirit of or against the guiding principles, we see people in the chats post the GIFs in and say 'great, yes, together we build [C7], you exemplified it well' or 'here's the reminder [for the guiding principle] focus.' It works very well, and it's always visible because of that. [...] and I honestly don't know how to get that into the real, physical communication world." (F7)

A particular advantage of using instant messengers is regarded as the fact that direct and situational reactions are made to messages (e.g., through GIFs and emojis), and interaction between employees is facilitated (F3; F7; F12).

Only a tiny proportion of communication still takes place via email in the start-ups (F2; F3; F10; M13). These are primarily important announcements or communication with external parties (M4; M6; M13). In start-up C14, emails are no longer used at all for internal communication: "*Internally, emails are banned, that's technologically totally outdated for internal communication.*" (F14b)

The communication channels and tools are adapted according to the size of the start-up, which in turn influences the corporate culture (F2). In general, it was emphasized that to maintain the corporate culture, it is essential to ensure structured communication (M9; F14a) across the board through the use of several communication channels and tools so that the relevant information is transported, and no information deficits arise (F2; M10; M11; M13; F12), e.g., All-hands meetings and Slack. The CPO of M13 underlined that consistent storytelling is required for all chosen communication channels and that internal communication should not deviate from external communication.

4.10. Active Participation and Involvement of Employees

Finally, another lever for the development of a positive and strong corporate culture was seen in the active and con-

scious participation and involvement of (all) company members in the activities to build or maintain the culture. This is accompanied by (27) **decision-making and development activities** regarding the start-up's culture.

For many respondents, corporate culture is not something that the founders solely predefine, even if they lay the initial foundations for it (F1; F7; F14b); instead, it is a collective, ongoing creation. For this reason, employees from various interviewed start-ups were involved in actively and consciously shaping it at an early stage (F1; F3; F5; M6; M9; M11; F12). The process of defining and evaluating corporate values together with the start-ups' employees has already been addressed in Chapter 4.9.

Further, employees are encouraged to contribute their ideas for corporate events introduced in a previous chapter and plan them accordingly (F1; F3; F10; F12):

> "To make the corporate culture tangible, we do a lot about events, and we pass the ball back and say, 'Hey, what would you like to have here as an event, and what should it be like? And do it yourself.' In other words, empower the people themselves to take the reins for it." (F1)

Beyond a specific company size, some start-ups have institutionalized cultural participation in the form of culture clubs (M6), committees (M9; M11), normative circles (M8), working groups (M9), or the like:

> "We have an open participation culture. That means, on the one hand, you can join the culture club. If you simply say, I want to participate, you can join. [...] But the fact that this became a little more institutionalized with the culture club came about later." (M6)

> "[The strategic off-site was] the trigger for many working groups that (...) [dealt with] how can we live [the corporate values] even better? So exactly these behavioral implications, (...) and then somehow such initiatives have developed." (M9)[16]

Participation in these is voluntary, and there are no direct extrinsic benefits for these employees who are members of such (M6). Additionally, above a specific company size, structures in the form of such institutionalized groupings for contributing ideas are seen as necessary to avoid duplication of effort (M9).

The view that every employee is a culture carrier and potential culture shaper, but above all, the founders and managers among them are the most influential, was shared by most interviewees (F1; M6; F7; M9; M10; M11; F14b). This observation has already been partly addressed in Chapter 4.1. However, in the founders' view, the majority of the first employees (approx. 30-50) are especially strong (28) **culture multipliers** by their nature and as they have internalized the companies' culture for several years, which is why these - along with other factors - are also developed further in management positions (cf. Chapter 4.2):

> "I think, [corporate culture] is very much communicated by those who have been there for a long time, (...) [who are] absolute culture carriers." (F3)

> "The early joiners, in particular, are influential leaders in the culture. The first - I would say - 50 employees. (...) especially the people in leadership positions, have a powerful influence on how corporate culture actually is." (M6)

Besides the founders and early joiners, managers are generally seen as further multipliers of the corporate culture, given their position and role model function, regardless of how long they have belonged to the start-up (F2; F3; M6; M13; F14b).

These multipliers are needed because, above a specific size, it is no longer feasible to pass on the corporate culture directly to all employees oneself, argued the co-founder of C5.

In addition to these culture carriers who occupy their role rather implicitly, there are also appointed (29) **culture ambassadors**. These have primarily been elected and have the intrinsic need to address the corporate culture thematically, regardless of the length of time they have been with the start-up or their position:

> "[The Ambassadors] are responsible for carrying the values into the company. (...) And it wasn't that we said we needed five people to be ambassadors, but rather who was up for it. And then we took precisely those people [that were elected into the culture team]. And for them, it is an intrinsic need to live by these [cultural] things and spread them further, so they don't get lost. Thus, they make sure, for example, that [corporate culture] is dealt with in onboarding [...]." (F7)

Employees who are part of a culture club or committee may be considered culture ambassadors, even if they do not have an official designation internally. This is based on the assumption that they also want to proactively participate in shaping and maintaining the corporate culture out of an intrinsic need.

Half of the respondents mentioned (30) **employee surveys** as another way to engage their employees in developing the start-up's culture (F1; F2; F5; M10; F12; M13; F14b). Those surveys are used to obtain input and feedback systematically, continuously, and anonymously from employees at

[16] This kind of translation work, translating core values into concrete behaviors and reflecting them in the respective development phase a start-up is currently in, emerged from several conversations. For example, the most important core value at start-up C8 is to protect the environment, which is why there is a "no flight" or "no print" policy, or only vegetarian catering.

certain intervals regarding company issues, but also to monitor the development of their culture. The survey intervals are decreased as the number of employees grows. This need to establish such a process was seen as a means of staying close to employees in a certain way and capturing every voice despite the increasing size of the workforce and the anonymity that goes with it (F12; M13; F14b). Formerly, the founders could gauge the workforce's mood by simply walking around the office and talking to a few individuals (M6; F12).

The questions in those surveys can directly or indirectly concern corporate culture. Various survey methods were mentioned in this regard: In addition to quarterly to annual company-wide and team-based surveys, through which employee satisfaction (e.g., eNPS - employee net promoter score), motivation, and engagement are to be captured via numerous questions, weekly to monthly pulse checks are used as well. Pulse checks are shorter employee surveys that measure ad hoc employee satisfaction or motivation and usually include repetitive questions. These surveys aim to listen to employees actively and measure cultural aspects, for example, how well the corporate values are lived from the employees' perspective and whether there are cultural differences between the teams or departments (M4; M9). By collecting the data, trends can be identified early, and, if necessary, appropriate measures and adjustments derived (F2; F14b). In addition to personal feedback, this type of feedback is also helpful in discovering barriers to growth and opportunities that may be perceived by employees (F5; F10; M8).

Besides internal data, external data is also collected to obtain a quantitative status quo picture of the start-up's culture. These sources include kununu.com and glassdoor.de, both websites on which companies are anonymously rated by their former or current employees. One of the categories to be evaluated is "corporate culture" or "corporate culture and values".

> "[It's important] that people can actively give feedback, in anonymized or non-anonymized form, depending on how they feel comfortable. [...] And [with such rapid growth], of course, you have to question things. We're just not 20 people anymore. With 250 people, we are a bit more anonymous. And you really have to take that on board and think about it very actively: What does the topic of ownership mean when you're 250 people vs. 20 people? And that is simply an ongoing process that we keep going. In that sense, each individual has a voice, and it's read. Kununu, glassdoor, name them all - we use all the data points to inform this topic of the corporate culture. [...] We also read these [comments on the rating portals] very actively and want to act accordingly." (F14b)

5. Discussion

The empirical study addresses the question of how start-ups develop and maintain their corporate culture even when they grow strongly in terms of personnel. For this purpose, in addition to analyzing existing literature, sixteen interviews with founders and managers were conducted, analyzed, and interpreted using qualitative content analysis. From the data, 30 instruments emerged, which relate to the cultural development and maintenance by the interviewees from diverse start-ups.

In the following, the insights gained through the interviews are summarized and interpreted holistically in Chapter 5.1, which focuses on the development and maintenance of a start-up's culture in the growth phase. This is followed by a reflection on the twelve mechanisms according to Schein and a discussion of the instruments that the founders and managers considered particularly relevant for maintaining their corporate culture despite rapid growth. Subsequently, the findings are placed in the context of the current research in Chapter 5.2. Finally, the limitations associated with this work are mentioned in Chapter 5.3, and possible future research fields are pointed out in Chapter 5.4.

5.1. Developing and Maintaining a Start-up's Culture - Summary of Key Insights

Application of Instruments in the Different Life Stages

Literature has shown that the formative phase of a company is primarily shaped by the founding team, which already brings and communicates an (implicit) set of values to the company through its behavior (Sackmann 2017, p. 315; Schein 1983). Another pillar of corporate culture is seen by some interviewees as the company's reason for existence (mission) together with its vision and values, which form the DNA of the culture (Jarnagin & Slocum, 2007, p. 292). They translate and disseminate the founding team's basic assumptions (Zheng et al., 2009, p. 160). According to these, the founders and management must live correspondingly as role models and align their behavior and decisions with the corporate values since their behavior is to be valued as symbolic actions (O'Reilly, 1989, p. 20). The importance of the role model function was also emphasized several times by the founders. Thus, the aspects labeled in this paper under Basic Entrepreneurial Orientation and the Founders' Behavior (Chapter 4.1) can be attributed to the cultural mode of *inspiration* of Zheng et al. (2009, p. 292) and shape the basic cultural direction. However, the interviewees regarded the Basic Entrepreneurial Orientation as a unifying force and the role of the founders as having a significant function in maintaining corporate culture in the growth phase even beyond the founding stage. As a result, it is still considered essential to foster particular points of contact with the founders to maintain the integrity of the company culture. These touchpoints are initiated during recruitment, onboarding, all-hands meet-

ings, and events, and thus in a staged environment (Schein, 2004, p. 258).

As start-ups progress into the growth phase, the corporate culture also evolves, which is why a different cultural mode becomes relevant (Zheng et al., 2009, p. 158). Based on the findings, it appeared that founders are confronted with the issue of their own corporate culture as soon as the first employees beyond the founding team join the company (F2; F3; F5; F7; M8; M9). Even if the culture was initially lived somewhat unconsciously and not actively shaped, but through interpersonal interactions, it became more critical with each employee beyond the founding team not to leave culture shaping to chance, and corporate values hung on walls but to consciously shape it (M10; M11). This implied utilizing instruments that supported the communication and implementation of the corporate values so that these become apparent and reinforced in everyday business (Jarnagin & Slocum, 2007; Schein, 2004).

An initial step in most start-ups was to write down the core values, making them directly accessible to employees. Even in start-ups such as C4 and C5, where corporate culture was deprioritized over other corporate issues in the early years, the feeling arose after a specific company size. It is essential to actively and consciously deal with their culture and its further development and thus to attach greater importance to it for the company's continued success (M4; F5). The founder of F5 had even expressed concerns that an earlier examination of the topic might have been necessary to maintain corporate culture in the desired direction. It can be concluded that addressing one's own corporate culture as early as possible can only be beneficial.

To achieve a more conscious and active shaping of the corporate culture and not to leave it to uncontrolled or even undesirable proliferation, as the number of employees increases rapidly, further instruments were selected and introduced to reinforce the start-ups' corporate values. This was accompanied with the establishment of processes and structures. It also goes hand in hand with the findings of the growth phase of Zheng et al. (2009, pp. 161-162) as well as Weeks and Galunic (2003, p. 1337), who assume that founders and corporate values alone cannot ensure that cultural modes of thought are effectively disseminated within the company; instead, they need to make use of other instruments accordingly. This was also evident in the start-ups, where more time and human resources were devoted to further shaping the corporate culture, and responsibilities were assigned to operationalize the corporate values.[17] For example, corporate values were integrated into processes such as recruiting, onboarding, or performance assessment and rituals such as giving awards for exemplary cultural behavior. This process by which cultural assumptions and values are embedded through "organizational systems, structures, policies, rites and rituals, stories, and other tangible forms" is what Zheng et al. (2009, p. 161) call the *implantation* mode. Due to their start-up's hyper-growth, interviewees pointed out that corporate culture is currently one of the top issues for the management team (M4; F5; F14b) and that there is a need to respond quickly to changes in the corporate environment by adapting processes and structures (M9; M13).

Although founders and managers significantly influence corporate culture and initiate appropriate measures to preserve their cultural mode of thought, the opinion prevailed that corporate culture cannot be prescribed purely top-down but should be co-designed by its employees. Interaction between the company members was seen as an essential and crucial factor in developing and maintaining the corporate culture. This was demonstrated by the initiation of numerous and regular meetings and gatherings of employees in various formats and the creation of corresponding platforms for active and conscious co-creation. Culture clubs or working groups dedicated to corporate culture were mentioned as examples. Instruments and formats that promote dialog between management and employees (e.g., feedback, discourse on corporate culture in onboarding) also gave the impression that the claim is not to carve corporate culture in stone but instead to capture voices from within the company and, if necessary, make adjustments in the implementation and further development of corporate culture. By actively involving employees, the prevailing culture may also be perceived as corporate culture and not as the founders' culture, which leads to consensus and a culture employees want to nurture and maintain together.

Further, successfully developing and maintaining a corporate culture is about "walking the talk". That is, translating corporate values into artifacts and actions visible to employees and deriving appropriate measures from them. One of C12's corporate values was "We believe in the power of health". Accordingly, sports classes were also offered to employees (F12). Such translation work gives authenticity to the corporate culture and creates a framework for living the corporate values. The realization of the start-ups' values and reinforcement of them in artifacts can be seen as the "proactive realization and symbolization" of the *implantation* mode, according to Zheng et al. (2009, p. 162).

At the same time, the research results showed that "talking the talk" was also practiced to a great extent in the start-ups studied, i.e., repeatedly pointing out the importance of the corporate culture. In weekly all-hands meetings, for example, time was allocated to talk about corporate values. Frequently talking about them may only have a reinforcing effect if "walking the talk" can already be observed. Therefore, the example of Personio's Culture Week can only have an impact if the majority of the corporate culture is already being lived and the values are also translated into action as part of this event. In this way, corporate culture can be experienced by the participants and is not just talked about and preached. Under this premise, an event in the name of corporate culture might indeed be culture reinforcing.

The study also showed that rapid growth creates a need

[17] In an experiment, Brown (2018, p. 190) found that only about ten percent of the organizations studied have operationalized their values into teachable and observable behaviors to train employees and influence accountability.

among the founders to express corporate culture in numbers and monitor it somehow, as the points of interaction between the founders and the numerous employees were no longer possible to the same extent as in the founding phase. One instrument mentioned for this purpose was the performance assessment, which evaluates how much an employee has delivered on the company's values. In addition, company surveys are conducted at regular intervals, which were considered as a trend indicator for changes in the corporate culture.

Schein's Twelve Mechanisms - the Right Framework for Start-ups?

In the following, the findings are related to Schein's twelve mechanisms and reflected upon.

First, not every single instrument can be assigned to just one mechanism, which is partly because some of the mechanisms cannot be clearly distinguished from each other or leave room for interpretation, such as "What leaders pay attention to", since it "can mean anything" (Schein, 2004, p. 247), on which leaders systematically pay attention to.

Further, Schein derived his findings on the twelve mechanisms by which founders and managers embed and reinforce corporate culture, consciously and unconsciously, based on a clinical study (Schein, 2004, p. 60). Thereby, drawing on his observations in various companies, he identified areas of influence on culture rather than practicable instruments from a founder's or manager's perspective.

Besides, Schein's research focused on the underlying basic assumptions, which tend to be unconscious, taken for granted, and difficult to grasp. Due to the characteristics of these basic underlying assumptions and the chosen research design, the focus of the data obtained in this study was rather on the articulated corporate values, which were seen by the interviewees as the basis for culture work, in addition to the observable behavior of management. From the data collected, it became apparent that, due to their rapid growth, start-ups strive to communicate very clearly to their employees how their corporate culture is to be understood and are less likely to rely solely on implicit actions that require initial deciphering. This can be explained by the fact that in many instruments mentioned, the corporate values have explicitly been implemented in several procedures.

The assumption that the observable behavior of founders and managers is seen as the most important instrument for culture development rather than documented cultural target statements or how offices are designed could be confirmed. It was repeatedly emphasized by the interviewees how important it was to have the right managers acting in the founders' interests and that all other instruments and measures must be consistent and present a uniform picture in this regard. This is in line with Schein's theory (2004, pp. 262-263).

Besides, Schein states that the primary mechanisms transmit clear messages, whereas the secondary mechanisms are ambiguous, less influential, and serve to strengthen the culture when they are consistent with the primary mechanisms. The latter become primary mechanisms only when the organization matures and stabilizes. However, this study has shown that some processes (e.g., onboarding, performance assessment, training, and workshops) that can be attributed to Schein's "organizational systems and procedures" convey clear messages in practice. The same holds true for stories shared in onboarding and assigning value awards. Moreover, the start-ups studied pay attention to the use of secondary mechanisms early in the growth phase. To establish a strong culture, this research recommends that these mechanisms should not be regarded as secondary but instead be applied as early as possible.

According to Schein, the socialization process of new employees is primarily embedded in daily work routines, which is why he does not perceive the need to teach new joiners about critical cultural assumptions explicitly (Schein, 2004, p. 262). The emphasis placed on the initial onboarding and training of new employees by interviewees challenges this assumption, possibly due to the rapid growth of start-ups and the need for explicit communication about their corporate culture.

Looking at Schein's instruments gives further the impression that shaping and influencing corporate culture rests solely with the founders and (later) the managers. The significance of employee co-creation is largely neglected in Schein's instruments. The interviews indicated that, from the founders' perspective, culture is formed top-down to a certain extent, and the impact gradient on it also runs accordingly. Nevertheless, the qualitative research underlined the importance of actively involving employees in the co-design and development of a strong corporate culture in start-ups. In this respect, the participative aspect should also be reflected in several instruments. Furthermore, the way start-ups communicate (open and transparent) and the explicit positioning of corporate culture in communication (e.g., meetings, onboarding, feedback) were attributed high significance in the interviews for developing and maintaining corporate culture. Therefore, the mechanism "Formal Statements" of Schein was subordinated to the instrument of "Communication" in this study.

The aspects listed above show that Schein's mechanisms may not be framed for the context of fast-growing start-ups, but rather for corporates growing steadily and slowly in a less volatile environment. Therefore, more research from the start-up and management perspective is needed to provide founders and managers with instruments to support corporate culture development and maintenance during stages of fast growth.

Key Instruments for Developing and Maintaining a Corporate Culture

Finally, the three instruments that the founders and managers considered most significant in the development and maintenance of the start-ups' culture - apart from their role model function are briefly examined.

The founders and managers saw the corporate *values* as the core and constant of the corporate culture, whether writ-

ten down or not. The fact that corporate values represent the central element of corporate culture is also reflected in literature (cf. Hofstede et al. 1990; Schein 2004; Trice and Beyer 1984).

The majority of interviewees considered the formulation and codification of core values particularly helpful, as it was regarded as an integral part of their corporate culture and governance. Parts of the founders' basic assumptions were explicitly expressed through an initial formulation, reducing room for interpretation and making the corporate culture's postulated core accessible to employees. They determined what every employee should strive for, how to interact with each other, and how to conduct business. In addition, it became apparent that these serve as a reference point in introducing and elaborating further cultural instruments to foster corporate culture, such as onboarding, performance assessment, or employee surveys. Thus, formulating the values is incumbent on a reinforcing mechanism to the extent that they are consistent with lived values (Schein, 2004, p. 262).

Even though the corporate values of the start-ups have hardly been changed over the further course of time, it tends to be considered essential to reflect on them in the context of experienceability during the various growth phases of the start-ups and to adapt or introduce instruments of cultural development and maintenance accordingly.

Founders and managers placed significant importance on *recruiting* to influence the desired corporate culture. Recruiting has already been discussed frequently in existing literature as a central instrument for culture development (cf. Chatman and Cha (2003), Sackmann (2017), Wiener (1988), and Willcoxson and Millett (2000)). Since the founders had direct influence over whom they hired into the company, this was considered as an essential driver in shaping, developing, and maintaining their culture. Based on the findings, most founders and managers hired and selected candidates who appeared to fit their culture, reinforcing certain aspects of the existing culture (Chatman & Cha, 2003, p. 26). Contrary to O'Reilly et al. (1991), start-ups tried to pursue this systematically and deliberately at a very early stage to reinforce the corporate culture. This is sought through an extensive application process, including several organizational members and evaluation criteria to test whether an applicant fits the start-up's culture. In start-ups, the founders still assessed the culture fit of applicants, which they claimed has a higher priority than technical fit. However, this was usually no longer possible with rapid personnel growth due to other obligations of the founders.

From the interviews, it emerged that the first employees made a significant contribution to preserving the corporate culture during rapid growth, as they are implicit culture carriers and further transmit the corporate culture into the company in the spirit of the founders; also, when introducing a management level from the role of a manager. In the case of rapid growth, it is therefore not only a matter of attracting new employees but also of retaining and leveraging existing ones in the company; otherwise, critical cultural multipliers of the founders may be lost. This is characterized by personal development measures and expression of appreciation towards employees (e.g., promotion, feedback, rewards, etc.).

Lastly, several cultural scholars consider corporate *rituals and events* central to transmitting cultural assumptions and maintaining established value systems (Schein 2004, pp. 266-267; Trice and Beyer 1984, pp. 654-655; Wiener 1988, p. 543). Numerous interviewees also affirmed this, who increasingly referred to the events and informal gatherings that took place regularly and intended to strengthen team cohesion and a sense of connectedness, enabling exchange and experiencing cultural values (O'Reilly 1989, p. 20; Trice and Beyer 1984, p. 657). The resulting interpersonal relationships strengthened the existing corporate culture and integrate new members (Willcoxson & Millett, 2000, p. 97). However, such events can be seen less as a control instrument for desired behavior, in contrast to performance assessment, for example, because they are subject to their own momentum.

5.2. Theoretical and Practical Implications

Firstly, this study is an exploratory attempt to investigate the crucial but largely unexplored topic of corporate culture, particularly the development and maintenance of it in a high-growth start-up context. The role of founders in the emergence and development of corporate cultures is already recognized by scholars (Sackmann 2017, p. 76; Schein 1983, p. 1; Schneider et al. 2013, pp. 371-372) and has also been emphasized in this paper several times. However, research on measures and practices for start-ups growing rapidly in terms of personnel that contribute to the development and maintenance of their culture is sparse. By focusing specifically on the instrumental level for culture development, a contribution is made to filling this research gap.

Secondly, previous research has suggested that different cultural modes exist in the founding and growth stages to address organizational needs (Zheng et al., 2009, p. 158). The analysis of the interview data revealed that during the founding phase, the vision, mission, and the founders' behavior are the primary culture-shaping elements (*inspiration mode*). Yet, this study has shown that the recruitment and selection of the first employees was also a key factor during the founding phase. In the (rapid) growth phase, start-ups dealt more actively and consciously with their corporate cultures (*implantation mode*). This was also reflected in the utilization of additional instruments introduced in this phase to mainly translate the start-ups' values into artifacts and desired behaviors and thus maintain the cultural core of the start-ups. Hence, the dynamic view on culture in this work is supported by this study (Weeks & Galunic, 2003, p. 1344) and that different cultural modes emerge at a start-up's founding and growth phase (Zheng et al., 2009, p. 158). However, this work has also shown that the inspiration mode continues to appear important in the growth phase, and the implantation mode is already considered by some start-ups when they have a relatively small workforce.

Thirdly, while the existing literature deals predominantly with generic instruments and perspectives for action for the management of corporate culture, this thesis has drawn from today's start-up context, which instruments have made a significant contribution to the development and maintenance of corporate culture and how they are implemented in practice. It became apparent from this study that the implementation of utilized instruments is continuously subject to an iteration process and adapted according to the current needs of the respective start-up (Does the format of the chosen instrument still fit, does it still fulfill its purpose with the current number of employees, are new formats needed if necessary?). Thus, this work has additionally shown that the instruments within the growth phase are subject to their own dynamics due to the rapid increase in the number of employees of the start-ups and the accompanying cultural needs of the company.

Not all start-ups studied applied all the instruments explored in this thesis. By considering the culture development instruments given in this paper, founders can intentionally and holistically develop a strong culture from day one, hence gaining efficiency advantages over entrepreneurs who only gradually realize the importance of culture. The practical implementation examples of the instruments provide start-ups with a more general orientation framework for developing and maintaining an effective culture, which, however, requires adaptation to individual needs. In this context, the core values should be the starting point for all development measures and the implementation of instruments.

5.3. Limitations

Even though initial implications for developing and maintaining a strong corporate culture in start-ups could be derived from the data in this thesis, the following limitations of the underlying methodological approach should also be considered when interpreting the results.

Firstly, qualitative research relies heavily on the researcher's subjective judgment and interpretation, especially in the cultural context. Researchers usually take an observational and interpretive role in studying corporate culture, which is why it cannot be studied without absolute objectivity (Schein, 2004, p. 51). The detailed category system tries to preserve the higher-level perspective and prevents the interpretive bias of the researcher to some extent. In addition, a detailed description of the research process (cf. Chapter 3) helps to make the procedure transparent to third parties and somehow reproducible through the utilization of the category system (Mayring, 2015, pp. 123-125). To further reduce intersubjectivity and promote discussion of perspectives, it is advisable to include at least a second researcher in subsequent studies (Gioia et al., 2013, p. 19).

Secondly, even though sixteen interviews provided rich answers to the research question and numerous culture development and maintenance instruments could be identified, the number of the sample is limited in its scope to draw generalized conclusions. Therefore, the instruments obtained should not be considered complete or definitive; instead, they should be seen as a decent selection to develop and maintain a strong corporate culture in fast-growing start-ups, obtained through a rather small sample of start-ups. Therefore, future research should consider a larger sample and add quantitative methods to validate the findings, which may strengthen the results or lead to further new insights and a more comprehensive and complex understanding of the chosen corporate culture instruments (Ehrhart et al., 2014, p. 143). In addition, it could be quantitatively verified what influence the collected instruments have individually on the development and maintenance of a strong corporate culture (Janićijević, 2011, pp. 83-84, 93).

Additionally, a prevalent shortcoming in interview situations is the recall bias, i.e., past events may not be remembered (correctly) at the time of the interview, or incidents may be given more or less importance in retrospect than originally (Beckett et al., 2001, p. 619). While some narratives refer to the present and short-term past, questions were also asked about the early days of the start-ups, making them more susceptible to retrospective bias. Hence, a more detailed analysis is needed to verify which key moments in culture development took place, when and with which intention, e.g., the writing down of core values respectively corporate principles or the introduction of a cultural institution such as a Culture Committee, which deal strategically with the start-up's culture. More insights can be gained from conducting a long-term study, i.e., from the time of founding to a particular growth in personnel, whereby several data points are collected (Ehrhart et al., 2014, p. 141). Thus, an evolutionary process can be illustrated more accurately, and insights into the dynamic view of culture in the context of personnel growth can be promoted.

Other limitations go hand in hand with the interview situation as such. Even though confidentiality and anonymity of what was said were indicated at the beginning of the interviews, it is possible that interviewees may not have revealed all aspects of the truth to an external researcher and may have concealed information they were reluctant to disclose or put themselves and their company in a positive light (Ehrhart et al. 2014, p. 292; Podsakoff and Organ 1986, p. 535). The actual observable behavior may therefore differ from the reported behavior (Baron & Hannan, 2002, p. 29). Thus, the extent to which corporate culture was actually considered in the introduction and design of all instruments such as team events, organizational structure, process design, etc., cannot be demonstrated. This could be counteracted by interviewing additional employees in critical positions.

Lastly, in this study, corporate culture was usually examined from the perspective of one founder or one manager who was among the first employees of the start-ups, which is why the narratives are mainly based on a person's testimony in this study. In two start-ups, data was collected from two people, whereby different insights could be generated. Depending on the functional orientation within one's own company, the interviewees may have unconsciously steered the focus towards a particular company area and thus instruments utilized there. Therefore, the founders' and managers' viewpoints are regarded as personal opinions on the

subject rather than generally accepted statements. In some instances, even contradictory statements were made. For this reason, it would be advisable to collect more data from various employees of the start-ups and include C-level, middle management to team members in the data collection to improve the validity of the results and to obtain a more comprehensive overview of different instruments and practices for developing and maintaining corporate cultures in start-ups (Podsakoff & Organ, 1986, p. 542).

5.4. Future Research

This study represents an initial exploratory approach to developing and maintaining corporate culture in start-ups at the instrumental level. Based on the results of this study and associated limitations, recommendations for further research can be derived as follows.

This study examined fast-growing start-ups that are considered culturally attractive, which is why further research is needed to also focus on start-ups that have lost attractiveness in terms of their corporate culture during rapid growth. Empirically collected data could yield further implications for the management of corporate culture and underline the importance of specific instruments and practices to avoid a negative trend in the corporate culture.

Further, hyper-growth start-ups with an average staff growth rate of more than 150 % over five years could be studied (Minola et al., 2015, p. 6). While the growth rates of some start-ups indicate hyper-growth, not all have reached a company age of five years. The extent to which the identified instruments for culture development and maintenance are effective in the long term due to further personnel growth, or the extent to which they are adapted, remains to be investigated. Besides, future research might focus on the impact of individual instruments on culture development and maintenance

Secondly, several start-ups studied grew especially rapidly during the global Covid-19 pandemic, in which businesses had to adapt to a fully remote mode of work intermediately. This study was not conducted with a focus on the impact of the global pandemic on corporate cultures (Spicer, 2020, p. 1738). Therefore, another area of research is the impact of the pandemic on the strength of a start-up's culture. Accordingly, it remains to be explored how implementing the instruments discussed in this study for developing and maintaining corporate cultures in start-ups will adapt again after the pandemic and in further growth. The pandemic also raises the question of how strong corporate cultures can be developed in start-ups that pursue a purely remote strategy and intend to do so in the future.

Since only the perspective of managers and founders was examined in this study, it would be beneficial also to survey a significant number of employees to examine which instruments are perceived by them at all and, from their point of view, have contributed significantly to the development and maintenance of their strong corporate culture. Including the perspectives of employees, the effectiveness of specific instruments such as events under a cultural motto, visits to other offices, employee surveys, or the awarding of value awards in all-hands meetings could be examined.

Finally, the topic of corporate values opened up another future research stream, as the interviewees also considered these to be an essential core of their corporate culture. However, as can be seen from the results, the stage of the start-ups varied in the stage of their lifecycles at which they had codified their corporate values. This can be a starting point for further research to determine if it has any impact on the company, whether the founding members already formulated the start-up's corporate values at the time of the founding or if they were the outcome of a company-wide process (few years) later. Research on this would provide further implications for management.

6. Conclusion

A great number of research results make apparent that developing a strong corporate culture can be essential to the success of companies and the well-being of their employees and conversely, that neglecting corporate culture can entail many costs for both companies and their employees (Warrick, 2017, p. 6) as demonstrated by the introductive example of N26. Start-ups, in particular, need a strong corporate culture to overcome the hurdles they face in their early and growth stages to prevail against competitors and to survive in the long run (Grossmann & Slotosch, 2015, p. 243).

There is a multitude of definitions of corporate culture and approaches to analyze, influence it consciously or unconsciously, or shape it purposefully. Furthermore, it was shown that corporate culture is not something static but evolves with the life cycles of a company. The culture of a start-up is no exception and is subject to even greater dynamics in its environment (Grossmann and Slotosch 2015, p.242; Ries 2011, p. 27). However, start-ups have hardly been the subject of research in the context of corporate culture.

The aim of this thesis was to show how start-ups approach the topic of corporate culture, develop, and maintain it during rapid growth. Therefore, a qualitative research approach was chosen to address the following research question: *How can the corporate culture of a start-up be positively developed and maintained while coping with a workforce growing significantly?* The focus of the research was on instruments that contributed to corporate culture development and maintenance in start-ups.

The findings indicate that start-ups, due to their rapid growth in personnel, rather see the need to create an environment in which the shaping and development of their corporate culture can take place consciously and actively. In addition to the vision and mission of the start-ups, active corporate culture formation starts with the definition of the company's values, which can be found as a common thread in various culture development and maintenance instruments as growth continues.

The larger the workforce gets, the more important it becomes to explicitly express the corporate culture and to introduce measures that allow the culture to be preserved at

its core, i.e., to maintain corporate values in the long term and to sustain them through the reflected use of appropriate instruments.

Even though the term corporate culture is a vague concept for many founders and managers that is difficult to put into words, this work has nevertheless shown that a large intersection of instruments used emerges among the various start-ups. Nevertheless, the instruments can only serve as a guiding framework, as the implementation of these differs in part among the start-ups and depends on their targeted corporate values.

Furthermore, it was shown that the implementation of the instruments is subject to a continuous adaptation process to adjust them according to the growing workforce of the company. At the same time, it became apparent that in addition to the role model function of the founders and managers, recruiting and various event formats but above all the defined corporate values, are assigned a significant contribution to the maintenance of corporate culture of start-ups in the growth phase.

By conducting a long-term study from the founding phase to the growth phase, a more accurate determination of the initial use and significance of the numerous instruments obtained in this study would be facilitated, as well as the requirements for these instruments that accompany rapid growth. Surveying employees of different ranks could also lead to further valuable insights.

Since corporate culture can be an essential factor for a company's competitiveness, especially for start-ups, this thesis contributed to creating awareness of it on a practical and scientific side, providing founders with a structured overview of instruments to develop and maintain their start-up's culture.

References

Adams, W. C. (2015). Conducting semi-structured interviews. In *Handbook of practical program evaluation* (4th ed., pp. 492–505).

Allan, J. S., Fairtlough, G., & Heinzen, B. (2002). *The power of the tale: Using narratives for organisational success*. Wiley.

Amazon. (2019). Unsere Leadership Principles. https://www.aboutamazon.de/news/arbeiten-bei-amazon/unsere-leadership-principles

BambooHR. (2022). *BambooHR: The Complete HR Software for People, Payrol & Benefits*. https://www.bamboohr.com/

Baron, J. N., & Hannan, M. T. (2002). Organizational blueprints for success in high-tech startups: Lessons from the Stanford project on emerging companies. *California Management Review*, *44*(3), 8–36.

Beckett, M., Da Vanzo, J., Sastry, N., Panis, C., & Peterson, C. (2001). The quality of retrospective data: An examination of long-term recall in a developing country. *Journal of Human Resources*, 593–625.

Bennis, W. (1986). Leaders and visions: Orchestrating the corporate culture. *Corporate culture and change*.

Brown, B. (2018). *Dare to lead: Brave work. Tough conversations. Whole hearts.* Random house.

Bryman, A. (1986). Leadership and Corporate Culture. *Management Decision*, *24*(6), 50–53. https://doi.org/10.1108/eb001426

Chatman, J. A., & Cha, S. E. (2003). Leading by leveraging culture. *California Management Review*, *45*(4), 20–34.

Crémer, J. (1993). Corporate culture and shared knowledge. *Industrial and Corporate Change*, *2*(3), 351–386.

Crosby, B. (2018). *How to maintain your culture in a rapid growth environment*. https://www.forbes.com/sites/forbeslacounci1/2018/08/17/how-to-maintain-yourculture-in-a-rapid-growth-environment/

Cummings, D. (2011). *The top 3 things every entrepreneur needs to know*. https://davidcummings.org/2011/12/07/the-top-3-things-every-entrepreneur-needs-toknow/

Deal, T. E., & Kennedy, A. A. (1982). *Corporate cultures*. Addison-Wesley.

Denison, D. (1990). *Corporate Culture and Organizational Effectiveness*. Wiley.

Denison, D. (1996). What is the difference between organizational culture and organizational climate? A native's point of view on a decade of paradigm wars. *Academy of Management Review*, *21*(3), 619–654.

Denison, D., & Mishra, A. (1995). Toward a theory of organizational culture and effectiveness. *Organization Science*, *6*(2), 204–223.

Dodge, H. R., & Robbins, J. E. (1992). An empirical investigation of the organizational life cycle model for small business development and survival. *Journal of Small Business Management*, *30*(1), 27–34.

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation*. Springer.

Drucker, P. (1991). Don't change culture use it. *The Wall Street Journal*.

Eberhardt, D. (2013). *Unternehmenskultur aktiv gestalten*. Springer.

Edwards, R., & Holland, J. (2013). *What is Qualitative Interviewing?* Bloomsbury.

Ehrhart, M. G., Schneider, B., & Macey, W. H. (2014). *Organizational Climate and Culture: An Introduction to Theory, Research, and Practice*. Routledge.

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

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, *50*(1), 25–32.

El-Namaki, M. (1992). Creating a corporate vision. *Long Range Planning*, *25*(6), 25–29.

Eurostat-OECD. (2007). Eurostat-OECD Manual on Business Demography Statistics.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational Research Methods*, *16*(1), 15–31.

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.

Glisson, C., & James, L. R. (2002). The cross-level effects of culture and climate in human service teams. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *23*(6), 767–794.

Gold, M. (2022). Remote Work: A Guide to the New Way of Working.

Gontard, M. (2002). *Unternehmenskultur und Organisationsklima: eine empirische Untersuchung* (Vol. 36). Hampp.

Graybill, J. O., Taesil Hudson Carpenter, M., Offord Jr, J., Piorun, M., & Shaffer, G. (2013). Employee onboarding: Identification of best practices in ACRL libraries. *Library Management*, *34*(3), 200–218.

Grossmann, C., & Slotosch, A. (2015). Die Unternehmenskultur in Startups - ideale Voraussetzung für den Einsatz von Social Media im Intranet? In W. Widuckel, K. De Molina, M. J. Ringlstetter, & D. Frey (Eds.), *Arbeitskultur 2020: Herausforderungen und Best Practices der Arbeitswelt der Zukunft* (pp. 239–252). Springer Fachmedien Wiesbaden.

Gründerszene. (2021). Hohe Personalfluktuation bei N26: „Es ist ein Mysterium, dass es irgendwie läuft". https://www.businessinsider.de/gruenderszene/fintech/die-hire-und-fire-politik-bein26-c/

Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, *18*(1), 59–82.

Guion, R. M. (1973). A note on organizational climate. *Organizational Behavior and Human Performance*, *9*(1), 120–125.

Hatch, M. J. (1993). The dynamics of organizational culture. *Academy of Management Review*, *18*(4), 657–693.

Hauser, F., Schubert, A., Aicher, M., Fischer, L., Wegera, K., Erne, C., & Böth, I. (2008). Unternehmenskultur, Arbeitsqualität und Mitarbeiterengagement in den Unternehmen in Deutschland: Ab-

schlussbericht Forschungsprojekt Nr. 18/05; ein Forschungsprojekt des Bundesministeriums für Arbeit und Soziales. https://www.ssoar.info/ssoar/handle/document/26571

Hoffman, R. (2017). Masters of Scale. https://mastersofscale.com/%5C#/reed-hastings-cultureshock/

Hofstede, G. (1984). *Culture's consequences: International differences in work-related values* (2nd, Vol. 5). Sage.

Hofstede, G. (1991). *Cultures and organizations: Intercultural cooperation and its importance for survival. Software of the Mind*. McGraw-Hill.

Hofstede, G., Neuijen, B., Ohayv, D. D., & Sanders, G. (1990). Measuring organizational cultures: A qualitative and quantitative study across twenty cases. *Administrative Science Quarterly*, *35*(2), 286–316.

Huber, G. P., & Power, D. J. (1985). Retrospective reports of strategic-level managers: Guidelines for increasing their accuracy. *Strategic Management Journal*, *6*(2), 171–180.

Hunter, J. S. (2021). N26 verliert Hunderte Mitarbeiter. https://financefwd.com/de/mitarbeiterschwund-n26/

James, L. R. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology*, *67*(2), 219–229.

Janićijević, N. (2011). Methodological approaches in the research of organizational culture. *Economic Annals*, *56*(189), 69–99.

Jarnagin, C., & Slocum, J. (2007). Creating corporate cultures through mythopoetic leadership. *SMU Cox School of Business Research Paper Series*, (07-004).

Jawahar, I., & McLaughlin, G. L. (2001). Toward a descriptive stakeholder theory: An organizational life cycle approach. *Academy of Management Review*, *26*(3), 397–414.

Jones, A. P., & James, L. R. (1979). Psychological climate: Dimensions and relationships of individual and aggregated work environment perceptions. *Organizational Behavior and Human Performance*, *23*(2), 201–250.

Kaschube, J. (1993). Betrachtung der Unternehmens-und Organisationskulturforschung aus (organisations-) psychologischer Sicht. In *Unternehmenskultur in Theorie und Praxis. Konzepte aus Ökonomie, Psychologie und Ethnologie* (pp. 90–146, Vol. 3). Frankfurt: Dierkes, Meinolf, von Rosenstiel, Lutz Steger, Ulrich.

Katila, R., Chen, E. L., & Piezunka, H. (2012). All the right moves: How entrepreneurial firms compete effectively. *Strategic Entrepreneurship Journal*, *6*(2), 116–132.

Kazanjian, R. K., & Drazin, R. (1989). An empirical test of a stage of growth progression model. *Management Science*, *35*(12), 1489–1503.

Kerr, J., & Slocum Jr., J. W. (2005). Managing corporate culture through reward systems. *Academy of Management Perspectives*, *19*(4), 130–138.

Kobi, J., & Wüthrich, H. (1986). *Unternehmenskultur: erfassen, verstehen, gestalten*. Landsberg: Moderne Industrie.

Kollmann, T., Kleine-Stegemann, L., Then-Bergh, C., Harr, M., Hirschfeld, A., Gilde, J., & Walk, V. (2021). Deutscher Startup Monitor 2021 - Nie war mehr möglich. https://startupverband.de/fileadmin/startupverband/mediaarchiv/research/dsm/dsm_2021.pdf

Kotter, J., & Heskett, J. (1992). *Corporate Culture and Performance*. New York: Free Press.

Kuckartz, U., & Rädiker, S. (2010). Computergestützte Analyse (CAQDAS). In *Handbuch qualitative Forschung in der Psychologie* (pp. 734–750). Springer.

kununu. (2022a). *N26 GmbH*. https://www.kununu.com/de/number-26

kununu. (2022b). *Was ist kununu?* https://www.kununu.com/at/kununu

Lourenco, R. (2021). *Yes, your company can hold on to during rapid growth*. https://www.toolbox.com/hr/culture/articles/yes-your-company-can-hold-on-to-itsculture-during-rapid-growth/

Martin, J. (2002). *Organizational research: Mapping the terrain*. Sage Publications.

Martin, J., & Powers, M. E. (1983). Truth or corporate propaganda: The value of a good war story. *Organizational symbolism*, *1*, 81–92.

Martinetti, C. (2020). *COO of C11: "Embrace Transparency In Order to Create a Fantastic Work Environment"*.

Mayring, P. (2015). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (12th ed.). Beltz.

Mayring, P. (2020). Design. In G. M. Mey & Katja (Eds.), *Handbuch qualitative Forschung in der Psychologie. Band 2: Designs und Verfahren* (2nd ed., pp. 225–237). Springer.

Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 633–648). Springer.

McGregor, L., & Doshi, N. (2015). How company culture shapes employee motivation. *Harvard Business Review*, *11*, 1–13.

Men, L. R., Chen, Z. F., & Ji, Y. G. (2018). Walking the talk: An exploratory examination of executive leadership communication at startups in China. *Journal of Public Relations Research*, *30*(1-2), 35–56.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2018). *Qualitative data analysis: A methods sourcebook* (4 (1994)). Sage publications.

Miller, D., & Friesen, P. H. (1984). A longitudinal study of the corporate life cycle. *Management science*, *30*(10), 1161–1183.

Minola, T., Cogliati, G., Cassia, L., & Paleari, S. (2015). Are hyper-growth firms inherently different? Preliminary evidence from a sample of fast-growing European SMEs. *International Journal of Entrepreneurial Venturing*, *1*.

Mintzberg, H. (1979). *The structuring of organizations*. Englewood Cliffs.

N26. (2022). *Über uns*. https://n26.com/de-de/ueber-n26

Neergaard, H., & Leitch, C. M. (2015). *Handbook of Qualitative Research Techniques and Analysis in Entrepreneurship*. Edward Elgar.

O'Reilly, C. A. (1989). Corporations, culture, and commitment: Motivation and social control in organizations. *California management review*, *31*(4), 9–25.

O'Reilly, C. A., Caldwell, D. F., Chatman, J. A., & Doerr, B. (2014). The promise and problems of organizational culture: CEO personality, culture, and firm performance. *Group & Organization Management*, *39*(6), 595–625.

O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of management journal*, *34*(3), 487–516.

O'Reilly, C. A., & Chatman, J. A. (1996). Culture as social control: Corporations, cults, and commitment. In *Research in organizational behavior: An annual series of analytical essays and critical reviews* (pp. 157–200, Vol. 18). Elsevier Science/JAI Press.

Ouimet, P., & Zarutskie, R. (2014). Who works for startups? The relation between firm age, employee age, and growth. *Journal of financial Economics*, *112*(3), 386–407.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2013). 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*(5), 533–544.

Patton, M. Q. (2002). Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective. *Qualitative Social Work*, *1*(3), 261–283. https://journals.sagepub.com/doi/abs/10.1177/1473325002001003636

Pearce, J. A., & David, F. (1987). Corporate mission statements: The bottom line. *Academy of Management Perspectives*, *1*(2), 109–115.

Personio. (2021). *#Accw: personio's first-ever all company culture week*. https://www.personio.com/blog/people-personio-all-company-culture-week/

Pettigrew, A. (1990). Organizational climate and culture: Two constructs in search of a role.

Picken, J. C. (2017). From startup to scalable enterprise: Laying the foundation. *Business Horizons*, *60*(5), 587–595.

Podsakoff, P. M., & Organ, D. W. (1986). Self-reports in organizational research: Problems and prospects. *Journal of management*, *12*(4), 531–544.

Reypens, C., Rückert, D., & Delanote, J. (2020). From starting to scaling: How to foster startup growth in Europe.

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Currency.

Rubin, H. J., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data*. Sage.

Sackmann, S. A. (2017). *Unternehmenskultur: Erkennen - Entwickeln - Verändern. Erfolgreich durch kulturbewusstes Management* (2nd ed.). Luchterhand: Springer.

Sathe, V. (1983). Implications of corporate culture: A manager's guide to action. *Organizational dynamics*, *12*(2), 5–23.

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.
Schein, E. H. (1983). The role of the founder in creating organizational culture. *Organizational dynamics*, *12*(1), 13–28.
Schein, E. H. (1992). *Unternehmenskultur: Ein Handbuch für Führungskräfte*. Campus-Verlag.
Schein, E. H. (2004). *Organizational culture and leadership* (3rd ed.). Jossey-Bass.
Schmitt, C. (2018). *Der lange Weg zum Erfolg: 6 Management-Erkenntnisse für aufstrebende Startups*. https://www.startupvalley.news/de/6-managementerkenntnisse-fuer-aufstrebende-startups/
Schneider, B., Ehrhart, M. G., & Macey, W. H. (2013). Organizational climate and culture. *Annual review of psychology*, *64*, 361–388.
Schultz, M. (1995). *On studying organizational cultures*. Walter de Gruyter.
Sørensen, J. B. (2002). The strength of corporate culture and the reliability of firm performance. *Administrative science quarterly*, *47*(1), 70–91.
Spicer, A. (2020). Organizational culture and COVID-19. *Journal of Management Studies*, *57*(8), 1737–1740.
Staehle, W. H. (1999). *Management. Eine verhaltenswissenschaftliche Perspektive* (8th ed.). Vahlen.
Steele, F. (1973). *Physical settings and organization development*. Addison Wesley Publishing Company.
Strauss, A. (1987). *Qualitative Analysis for Social Scientists*. Cambridge University Press.
Strauss, A., & Corbin, J. (1998). *Basics of Qualitative Research. Techniques and Procedures for Developing Grounded Theory* (2nd ed.). Sage Publications.
Trice, H., & Beyer, J. M. (1984). Studying organizational cultures through rites and ceremonials. *Academy of Management Review*, *9*(4), 653–669.
Trice, H., & Beyer, J. M. (1993). *The cultures of work organizations*. Prentice-Hall, Inc.
Tsui, A. S., Zhang, Z.-X., Wang, H., Xin, K. R., & Wu, J. B. (2006). Unpacking the relationship between CEO leadership behavior and organizational culture. *The Leadership Quarterly*, *17*(2), 113–137.
Valencia, J. (2019). Scaling Culture in Fast-Growing Companies. *Harvard Business Review*. https://hbr.org/2019/06/scaling-culture-in-fast-growing-companies
Van Vianen, A. E. (2000). Person-Organization Fit: the Match between Newcomers' and Recruiters' Preferences for Organizational Cultures. *Personnel Psychology*, *53*(1), 113–149.
Warrick, D. D. (2017). What leaders need to know about organizational culture. *Business Horizons*, *60*(3), 395–404.
Waterman, R. H., & Peters, T. J. (1982). *In search of excellence: Lessons from America's best-run companies*. Harper & Row.
Weeks, J., & Galunic, C. (2003). A theory of the cultural evolution of the firm: The intraorganizational ecology of memes. *Organization Studies*, *24*(8), 1309–1352.
Welch, C., & Piekkari, R. (2006). Crossing language boundaries: Qualitative interviewing in international business. *Management International Review*, *46*(4), 417–437.
Wiener, Y. (1988). Forms of value systems: Focus on organizational effectiveness and cultural change and maintenance. *Academy of management review*, *13*(4), 534–545.
Willcoxson, L., & Millett, B. (2000). The management of organisational culture. *Australian Journal of Management and Organisational Behaviour*, *3*(2), 91–99.
Witt, H. (2001). Forschungsstrategien bei quantitativer und qualitativer Sozialforschung. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *2*(1), 9.
Witzel, A. (2000). The Problem-centered Interview. *Forum Qualitative Sozialforschung*, *1*. https://doi.org/https://doi.org/10.17169/fqs-1.1.1132
Yin, R. K. (2009). *Case Study Research: Design and Methods* (4th ed.). Sage.
Yin, R. K. (2018). *Case study research and applications* (6th ed.). Sage.
Zacharakis, Z. (2020). „Gegen fast alle Werte, an die wir bei N26 glauben". *Zeit Online*. https://www.zeit.de/wirtschaft/unternehmen/2020-08/betriebsrat-n26-arbeitsbedingungen-einstweilige-verfuegungmanagement/komplettansicht?utm_referrer=https%3A%2F%2Fwww.iphoneticker.de%2F
Zheng, W., Qu, Q., & Yang, B. (2009). Toward a theory of organizational cultural evolution. *Human Resource Development Review*, *8*(2), 151–173.

Junior Management Science 9(3) (2024) 1769-1804

Junior Management Science

www.jums.academy
ISSN: 2942-1861

How Does ESG Rating Disagreement Influence Analyst Forecast Dispersion?

Robin Spira

University of Bayreuth

Abstract

The practice of responsible and sustainable investing has led to the incorporation of environmental, social and governance (ESG) information into investment decisions. The role of ESG rating agencies has been to facilitate decision-making by aggregating unstructured ESG information into a single rating. Market participants, such as financial analysts, rely on these ratings as part of their research. However, ESG rating agencies rarely agree in their assessment of a company's ESG performance, leading to divergent ESG ratings. This paper uses an OLS regression model based on a large sample of firm data to investigate whether ESG rating agency disagreement increases analysts' forecast dispersion. It builds on previous research by Kimbrough et al. (2022). The results do not provide sufficient evidence to support a significant relationship between ESG discrepancies and analyst forecast dispersion. This calls into question the importance of non-financial ESG information in analysts' assessment of a company's financial performance.

Keywords: analyst forecast; disagreement; ESG rating agencies; ESG score; intermediaries

1. Introduction

In the last ten years, the expanding practice of sustainable and responsible investing has resulted in the incorporation of environmental, social, and governance (ESG) information into investment decisions. An estimated US$ 35 trillion in assets under management are now invested with ESG information in mind (Global Sustainable Investment Alliance (GSIA), 2021, p. 9). Meanwhile, the parallel increase in demand from stakeholders for accurate information on firms' ESG performance, has led to the formation of ESG rating agencies. ESG rating agencies are third party information intermediaries that provide quantitative evaluations of a firm's ESG performance (Scalet & Kelly, 2010, p. 71). The concept of ESG performance intents to describe how well a firm manages its ESG risks and opportunities (MSCI, 2022b, p. 3). The final result of this evaluation is then compiled into an ESG rating score. In 2018 alone, investors spent $ 500 million on ESG ratings, highlighting their importance for guiding investment decisions (Gilbert, 2021).

However, there is considerable disagreement about what makes an investment sustainable and responsible. ESG rating agencies rarely agree in their assessment of a firm's ESG performance. This is remarkable considering how often credit rating agencies align in their assessment (Sindreu & Kent, 2018). Consequently, regulators and the media have raised concerns about whether ESG ratings can effectively guide investment decisions (Christensen et al., 2021, p. 147). If there is no agreement among rating agencies, ESG ratings might mislead market participants. Investors need to understand what the methodology chosen by ESG rating agencies actually measures and why. Otherwise, ESG ratings risk "creating a false sense of confidence among investors who don't really understand what lies behind the numbers – and therefore don't really understand what they're buying" (Allen, 2018).

One important group that relies on ESG ratings are financial analysts. Financial analysts are professionals who perform financial analyses on behalf of their clients to help them make investment decisions. To conduct those analyses, financial analysts use varous types of information about firms,

I want to thank Jan Seitz, who was my advisor at the University of Bayreuth, for giving me the opportunity to write my master thesis, always having great suggestions, providing new and great ideas to improve my thesis and generally supporting me throughout the time of writing.

DOI: https://doi.org/10.5282/jums/v9i3pp1769-1804

including ESG information (Wansleben, 2012, p. 407-410). Non-financial ESG-related information are valuable because they provide insights into firm-related risks and opportunities. But, due to the inconsistencies in the way different firms report ESG information and a lack of standardization, financial analysts increasingly rely on ESG rating agencies to analyze ESG information (Kotsantonis and Serafeim 2019, p. 53; Doyle 2018, p. 8 f.).

Over the past five years, there has been considerable progress in the literature on why ESG rating agencies disagree that much. For instance, research shows that scope, weighting and measurement (Berg et al., 2022, p. 1335 f.), the use of different data imputation methods (Kotsantonis & Serafeim, 2019, p. 54), and greater ESG disclosure (Christensen et al., 2021, p. 34 f.) lead to greater ESG disagreement.

Though, there are few studies that examine the impact of ESG rating disagreement on analysts' forecast dispersion. Previous studies have examined the relationship between credit ratings and analyst forecast dispersion (Avramov et al., 2009, p. 101), or the empirical association between CSR and information asymmetry (Cho et al., 2013, p. 81 f.). Dispersion is often interpreted as a measure of uncertainty and information asymmetry (Barron et al., 2010, p. 333). Other studies have examined how mandatory ESG disclosure affects the accuracy and dispersion of analysts' earnings forecasts (Krueger et al., 2021, p. 35), or the relationship between ESG disagreement and analyst forecast dispersion for US firms (Kimbrough et al., 2022, p. 29 f.). However, no study has yet examined the relationship between ESG disagreement and analyst forecast dispersion globally.

With this thesis, I attempt to fill this research gap by empirically investigating the influence of ESG rating disagreement on analyst forecast dispersion in an international setting. Forecast dispersion might reflect the amount of information commonly available to analysts (Han & Manry, 2000, p. 119). When analysts share a common forecasting model and observe the same firm-provided disclosures but have different private information, they will place less weight on their private information as the informativeness of firm-provided disclosure increases, decreasing forecast dispersion. The more ESG-related information a firm is disclosing, the lower the dispersion of analysts' earnings forecast should be (Lang & Lundholm, 1996, p. 471). Contrary, a high dispersion might suggest a lack of public information and hence analysts rely more on their own private information. Alternatively, greater dispersion could also indicate less agreement among analysts due to the inability or unwillingness of some analysts to fully and objectively gather and process ESG-related information (Behn et al., 2008, p. 330). If analysts have the same firm-provided and private information but put different weights on the components of firm-provided disclosure in forecasting earnings, additional disclosure may increase the dispersion of analyst forecasts (Lang & Lundholm, 1996, p. 471 f.). I predict that the dispersion of analysts' forecasts is not due to a lack of ESG-related disclosure, but rather due to discrepancies in the evaluation of ESG information. However, analysts often rely on ESG rating agencies to make sense of ESG-related information. Rating agencies that differ in the scope, weighting, and measurement of ESG-related information (Kotsantonis & Serafeim, 2019, p. 53). Consequently, the disagreement between ESG rating agencies should increase analysts' forecast dispersion.

The remainder of this thesis is structured as follows: Chapter two describes the characteristics of financial analysts and their practices. Chapter three focuses on the integration of ESG criteria into investment decision making. Afterwards, chapter four investigates the ESG rating agencies and their disagreement. Chapter five develops the hypothesis for the association between ESG rating disagreement and analyst forecast dispersion. After that, chapter six outlines the empirical study. Chapter seven and eight interpret the empirical results. Chapter nine highlights the limitations of this study and future research opportunities. Finally, chapter ten concludes.

2. Financial Analysts

2.1. Historical Background of the Profession

This chapter briefly introduces the reader to the emergence of financial analysts as a profession. Before the twentieth century the practice of finance was not yet associated with professional status. Only after that the profession of financial analyst emerged (Wansleben, 2012, p. 408). A defining moment for the financial analyst profession was the introduction of the stock ticker in 1867 (Preda, 2006, p. 754). Prior to its introduction, price information would be delivered by messengers and stocks may trade using numerous ticker symbols and sometimes even different prices. Thus, the stock ticker enabled market participants to monitor firm prices more efficiently (Fisher, 2019). With the introduction of the stock ticker, a subset of financial analysts known as technical analysts emerged. Technical analysis rests on the assumption of repetitive price behavior than can be analyzed by focusing on trends in stock prices (Wansleben, 2012, p. 418 f.). The other subset of financial analysts known as fundamental analysts emerged much later in the 1930s. Their predecessors were statisticians and accountants in banks, not technical analysts. The reason for the late appearance of fundamental analysts was that they encountered serious obstacles, as neither firms nor financial insiders shared information about corporate fundamentals before 1929 (Knorr-Cetina, 2011, p. 429). Although analysts had developed practices to interpret firms before the 1930s, they simply lacked reliable data. This changed with the 1933 and 1934 Act in the US. While the 1933 Act established laws for new issuances, including registration and disclosure requirements, the 1934 Act focused on annual, biennial, and event-related reporting requirements for traded firms (Benston, 1973, p. 133). The disclosed information allowed fundamental analysts to accurately interpret firms' earnings power and value (Jacobson, 1997, p. 25).

Equally important for the rise of the financial analyst profession was the ongoing financialization of the US economy

and public. During the 1950s share ownership doubled (Jacobson, 1997, p. 109). This surge in stock ownership not only created a demand for investment advice, but also fostered public legitimacy for analysts. This can be seen as a critical process in which financial analysts ultimately were in a position to ask questions and executives had to answer (Jacobson, 1997, p. 7). Another critical development has been the development of the certified financial analyst (CFA) examination as well as its worldwide acceptance. The standardized curriculum provided a source of legitimacy to the analyst practice. As Ketchum (1967, p. 35) points out, knowledge and its application builds the "keystone of a profession". (Wansleben, 2012, p. 411 f.)

The next chapter focuses on the practices of financial analysis and earnings forecasting commonly used by financial analysts. The purpose is to develop an understanding of how financial analysts evaluate the performance of firms and to show the reader what types of information are used in their evaluation.

2.2. Analyst Practices

2.2.1. Process of Financial Analysis

Collecting and Organizing Information

Financial analysis describes the process of collecting, processing, and evaluating fundamental information about firms and deriving investment recommendations for clients based on the analysis. Therefore, the first step is to collect and organize all relevant information about the firm.

The primary source of information is firm data. Such as financial statements, annual and quarterly result announcements, press releases, and other related news (Barker, 1998, p. 10). With these sources of information, though, analysts must always be cautious and question the reliability of the disclosed information. After all, firms are pursuing their own self-interest and may engage in creative accounting, window dressing or downright falsification of their books (B. Graham & Dodd, 2009, p. 68). Besides that, financial analysts attend analyst conferences, maintain intensive contact with investor relations representatives, and visit corporate headquarters and production facilities to fill the gaps left by disclosed firm information (Mars, 1998, p.86-111). In addition to firm information, analysts also draw on other sources of information for their analysis. In principle, any kind of information that can eventually affect future market developments is relevant. This can include all kinds of newspapers, business reports, books or studies, or other information sources on macroeconomic, political and social trends. In addition, personal contacts to sell-side analysts, external think tanks, firm representatives and people from academia as well as textbooks on financial analysis play an important role (Leins, 2018, p. 75-77). Hence, there is a wide range of financial information sources that analysts draw on.

In the last five years, non-financial ESG information has become an increasingly important source of information for analysts. According to the CFA Institute, 85% of their members now consider E, S, and/ or G factors when making investment decisions (CFA Institute, 2020, p. 4). This change is based on the view that integrating ESG factors into financial analysis allows for a more thorough assessment of both idiosyncratic and market-wide risk, as well as growth opportunities, which can improve long-term risk-adjusted returns (CFA Institute 2020, p. 27, MSCI 2022b, p. 2). Financial analysts draw from a mix of internal and external ESG information. On the one hand, they evaluate ESG information published directly by firms in their financial and statutory reporting. However, the consistency and comparability of ESG information from firms is poor because regulations on disclosure and reporting standards are still in development (CFA Institute, 2020, p. 37 f.). On the other hand, they draw on ESG ratings from rating agencies such as MSCI and Sustainalytics. 63% of financial analysts use them for their firm analysis. Still, a major problem with these ratings is that they vary widely across different rating providers. State Street Global Advisors reports a correlation of only 0.53 between the ratings of MSCI and Sustainalytics for firms in the MSCI World Index. These rating discrepancies result from differences in the collected data, conducted research, and models used to generate ratings, including valuation methodologies and weighting of various ESG information. (CFA Institute, 2020, p. 40)

Yet not all sources of information are equally valuable. Barker (1998, p. 11) surveyed analysts about their prioritized sources of information. He finds that personal contacts are particularly important to analysts (see Table 11 in the annex). By speaking to firm representatives, analysts seek to gain information advantages that goes beyond the disclosed information. These can be, for example, clarifications of financial statement notes, opinions on the firms' economic positioning relative to competitors or projections of next quarterly sales in a segment. Yet, the study does not include non-financial information. In addition to the source of information, there are four information attributes that matter to financial analysts. The information itself must be either timely, applicable, credible, or original to be of value (see Table 12 in the annex). First, the timeliness of the information matters. After financial analysts have analyzed a specific piece of information, and a widely accepted interpretation has taken hold among participants in the financial market, the data is deemed to be incorporated into the price. Consequently, the information loses its relevance for financial analysts. (Leins, 2018, p. 78 f.). Weekly newspapers, such as the Economist, serve as a good example. By the time the financial analyst reads the newspaper, the information has already been priced in for a few days. Consequently, weekly magazines are not really useful for the analyst in terms of the timeliness of their information (Leins, 2018, p. 80). Second, the applicability of the information also plays an important role. Applicability in this context means the usefulness of the information for the market forecasts. A highly applicable information often already contains information on how it could influence financial markets and links the informa-

tion to specific firms, economic sectors or market regions. This is very helpful because identifying the potential impact of information on financial markets is one of the most challenging tasks for financial analysts (Leins, 2018, p. 84 f.). An example for an applicable information source is Barron's. The magazine evaluates market trends and draws up implications for firms and industries. The third criteria financial analysts use when evaluating information is credibility. Credible sources help financial analysts in crafting inventive narratives while concurrently strengthening their position as experts in finance. Academic research, in particular, is often considered a highly credible source of information (Leins, 2018, p. 88). The fourth criteria is originality. Analysts can promote their forecasts as unique and inventive if they employ information that has not already been used by other analysts. An seeming unique market perspective gives investors the impression that they have been provided resources to help them navigate the uncertainties of financial markets. This is important because investors have many ways of assessing financial market data. Within this context, analysts must generate unique statements to capture their audience's interest. (Leins, 2018, p. 91-94).

Forecasting and Valuation

After collecting and organizing all relevant information, the next step for financial analysts is to make projections and to evaluate whether a firm is a good or a bad investment based on its current share price. For this, financial analysts need to evaluate a firm in terms of its underlying intrinsic value. According to B. Graham and Dodd (1934) this intrinsic value "is understood to be that value which is justified by the facts, e.g., the assets, earnings, dividends, definitive prospects, as distinct, let us say, from market quotations established by artificial manipulation or distorted by psychological excesses" (B. Graham & Dodd, 2009, p. 64). Financial analysts estimate the intrinsic value of a firm after evaluating all relevant information at their disposal. Investors can profit from their evaluation when the intrinsic value deviates from the market value of a firm. This occasionally happens because the price of the shares is based on what investors believe those shares are worth (Koller et al., 2020, p. 80). Having said that, financial analysis is by nature not an exact science (B. Graham & Dodd, 2009, p. 61). Financial analysts can only calculate the intrinsic value of a firm to the best of their ability and the knowledge available to them.

To calculate intrinsic value, financial analysts need to know how value is created. The concept of value has been introduced by Alfred Marshall in 1890 and has proven to be both lasting in its validity and difficult in its application. In short, the two main drivers of value are growth and return on invested capital (ROIC). Growth can be achieved either organically through general market expansion or by gaining relative market share, or inorganically through mergers and acquisitions (Koller et al., 2020, p. 260). ROIC, by contrast, is the result of a competitive advantage that allows the firm to either command premium prices or to enhance the efficiency of its production process (Koller et al., 2020, p. 224). Firms create value when they grow, and earn a ROIC greater than their opportunity cost of capital (Koller et al., 2020, p. 53). Firms that invest in revenue growth and improving their ROIC will generate higher discounted values of future cash flows. However, there is one caveat. Growth alone is not enough to realize higher discounted future cash flows (see Figure 2 in the annex). In cases where the return on capital is below the firm's cost of capital, higher growth actually leads to a reduction in the discounted value of future cash flows (Koller et al., 2020, p. 94 f.). Hence, firms should try to find the combination of revenue growth and ROIC that produces the highest discounted value of future cash flows.

Non-financial factors such as ESG can also be a value driver for firms. According to Henisz et al. (2019), ESG creates value in five ways. First, it facilitates revenue growth. Regulators are more inclined to grant access, permits and licenses to firms with a strong ESG position. Hence creating new opportunities for growth. Customers are also willing to pay and additional 5% for a green product. Second, ESG reduces costs. Among others, a strong ESG position can help to increase resource efficiency and thus reduce operating expenses such as raw-material costs and the true cost of water or carbon. Resource efficiency can boost operating profits as much as 60%. Third, ESG reduces regulatory and legal interventions. A strong ESG position can reduce a firm's risk of harmful state intervention. According to the study, one-third of corporate profits are at risk from state interventions. Fourth, a strong ESG position may boost employee productivity. It allows firms to attract and keep talented staff, boost employee motivation by providing them a sense of purpose and enhance overall productivity. Fifth, ESG can improve long-term returns on investment and capital allocation. For example, by allocating capital to more sustainable investment opportunities, which reduces the risk of future write downs and divestments (Henisz et al., 2019, p. 3-8).

The next step for financial analysts is to use one of various valuation methods to estimate the value of a firm. The most commonly used valuation method is the discounted cash flow (DCF) method. This method discounts future cash flows by the opportunity cost of capital. The idea behind it is that future cash flows are worth less because of the time value of money and the riskiness of future cash flows and thus need to be adjusted (Koller et al., 2020, p. 86). The discounted present value of future cash flows in this case represents the intrinsic value of the firm. By capturing the future performance of a firm in a single number, financial analysts can determine whether a firm is undervalued or overvalued relative to its market price. They can also compare different firms with each other. The traditional DCF method includes only financial numbers. But, non-financial ESG factors can be integrated into the DCF method with little effort. This is because ESG factors are often material and influence the firm's long-term cash flows (Wild, 2017, p. 54 f.). One shortcoming of the DCF method, however, is that each year's cash flow provides little information about the firm's competitive position and economic performance. Declining free cash flow

may indicate either poor performance or investment in the future (Koller et al., 2020, p. 305).

For the DCF method to work, financial analysts need to make projections about future cash flows. Yet, the further cash flows are in the future, the less accurate the projections become (Asquith & Weiss, 2016, p. 359). Graham and Dodd point out this problem in their book Security Analysis. They write, "some matters of vital significance, e.g., the determination of the future prospects of an enterprise, have received little space, because little of definite value can be said on the subject." (B. Graham & Dodd, 1934, p. vii). This leads to the problem of deciding how many years into the future to forecast and how detailed the forecast should be. Depending on the duration of the forecast, the financial analyst will arrive at different DCFs. In addition, there is also the problem of setting appropriate growth rates, interest rates, taxes, etc. Consequently, calculative approaches such as the DCF method can never produce precise results. They are always approximations of the future which are prone to errors (Leins, 2018, p. 72). To compensate for these uncertainties, some financial analysts create several cash flow scenarios (Winroth et al., 2010, p. 10). Others adjust their numbers according to the analyst consensus. Still others rely on their gut feeling or tweak the numbers to their liking (Leins 2018, p. 11 f. Wansleben 2012, p. 417 f.). As far as ESG factors are concerned, they usually have an impact over a longer period of time. Assessing ESG factors and their impact can therefore provide essential insights into future value drivers and thus improve long-term forecasting capabilities (Wild, 2017, p. 55 f.).

In the past, financial analysts and investors used earnings rather than DCF to calculate the intrinsic value of a firm. To use earnings as a measure of value creation is in principle not a bad idea, since firms that create value often also have attractive earnings and earnings growth. Moreover, earnings equals cash flow over the lifetime of the firm (Koller et al., 2020, p. 195 f.). However, practitioners have moved away from this method. The reason for this is that not all earnings create value. Margin improvements that come purely from cost cutting, e.g. research and marketing expenses, hurt value creating in the long term (Koller et al., 2020, p. 195). Furthermore, earnings can be accounting fiction (B. Graham & Dodd, 2009, p. xxx). Almost all firms need to invest in plant, equipment, or working capital. Free cash flow is what's left for investors once investments have been subtracted from earnings (Koller et al., 2020, p. 92). For simplicity, financial analysts and academics have sometimes assumed that all firms have the same ROIC. If this were the case, differences in the firms' cash flows would only result from differences in growth, making earnings growth a suitable measure of differentiation (Koller et al., 2020, p. 87 f.). Though, sometimes short-term earnings are the only reliable data available to financial analysts. In particular, when the uncertainty about the firm is so great that the cash flow cannot be accurately calculated. In this case, earnings are of great importance to the financial analyst (Koller et al., 2020, p. 204).

In addition to the DCF and earnings method, there are several other valuation methods worth mentioning. However, I will confine myself to valuation multiples and liquidation value, because I consider these to be the most important. Valuation multiples assume that similar assets should trade for a similar price. Firms in the same industry and with similar performance should trade at the same multiple. The most popular valuation multiple is the price-to-earnings (P/E) multiple, which is simply the equity value of the firm divided by its net income (Koller et al., 2020, p. 559). The advantage of these multiples is that they do not face the problem of inputs based on estimates, because only the market price and financial statements are needed for the calculation (Wansleben, 2012, p. 416). One major problem, nevertheless, is whether the firms are comparable at all. This requires a close look at the financial statements. For example, a firm with more debt relative to equity should trade at a lower P/E ratio than a firm with no debt, because more debt means higher risk for shareholders and a higher cost of equity (Koller et al., 2020, p. 559 f.). Also, comparisons of different ratios across different industries and among different firms might be misleading. Another problem is that the market valuation might be inflated by a speculative bubble or estimates of earnings, book value, and so forth can be wrong (Wansleben, 2012, p. 416 f.). Occasionally, DCF and valuation multiples may be inappropriate. This is the case, for instance, when the firm is expected to cease operations. Then it makes more sense to use the liquidation value (Asquith & Weiss, 2016, p. 354). The choice of the right valuation method therefore depends on the circumstances of the firm. In some cases, the use of several valuation methods may even have complementary benefits.

Investment Recommendation

After having determined the value of firm, financial analysts make investment recommendations to their clients based on their financial analysis. To underline their reports, financial analysts use persuasive charts, tables, and illustrations (Riles (2006, 2011) in Leins (2018, p. 12)). Analysts' recommendations are influential, as is evident from the changes in the price of a firm's stock after their release. Especially if the recommendations are widely publicized through the media or are issued by analysts with high credentials (Securities and Exchange Commission (SEC) 2010; Brown et al. 2009, p. 107). Ryan and Taffler (2004, p. 51) find that analyst activities such as issuing earnings forecasts and investment recommendations are associated with a 17% change in the market-adjusted price of stocks on the London Stock Exchange.

In general, one can distinguish between five different types of investment recommendations. These are Sell, Underperform, Hold, Buy and Strong Buy, whereby the in-between levels Underperform and Buy indicate a weaker conviction of the analyst. That said, not all recommendations carry the same weight. In fact, recommendation have to be assessed relative to the analyst's previous recommendation and the consensus opinion. For example, if a financial

analyst just reiterates the same rating, it carries less weight. Or, if the analyst merely issues a recommendation in line with the consensus. Contrary, if the financial analysts releases a recommendation out of line with consensus, it carries more weight, because the analyst stands aside from the safety of the herd and takes a greater reputational risk (Brown et al., 2009, p. 92).

Ultimately, however, it is the customer who decides what to do with the information. The analysts' report only provides information regarding the cost or benefit of investing in a certain stock. Whether the customer can ultimately profit from this information is, nevertheless, an open question. After all, financial analysts often have a conflict of interest when it comes to their recommendation. Customers should therefore critically scrutinize and compare the information (Securities and Exchange Commission (SEC), 2010). According to Winroth et al. (2010), the more sophisticated financial clients are more interested in discussing facts, underlying assumptions and arguments than in recommendations themselves. This is because institutional investors typically use information from several analysts, comparing their assessments (Winroth et al., 2010, p. 10 f.).

The next chapter focuses on the practice of earnings forecasting and forecast dispersion. The aim of this chapter is to show the reader the differences between financial analysis and earnings forecasts. It also aims to build a theoretical foundation for the dependent variable of this master thesis.

2.2.2. Practice of Earnings Forecasting

Forecast Estimates

Earnings forecasts are ubiquitous in today's financial markets. Investors rely heavily on earnings forecasts when making investment decisions (Givoly & Lakonishok, 1980, p. 221). Givoly and Lakonishok (1983) mention that „Earnings per share emerge form various studies as the single most important account variable in the eyes of the investors" (Givoly and Lakonishok (1983) in Jennings (1985, p. 1)). This view contradicts sharply with the notion that the value of a firm is equal to its discounted long-term cash flows. However, due to the unpredictability of future cash flows, practitioners use earnings as a reasonable proxy for DCF. Accounting earnings are well defined, and public firms' earnings statements are subject to thorough audits before they are published. As a result, investors consider earnings to be fairly reliable and convenient measure to value public firms (McClure, 2022).

The economic importance of earnings forecasts can also be seen in the amount of resources devoted to the preparation and analysis of such information by the investment community. Large brokerage firms employ large amounts of financial analysts to produce earnings forecasts. These sell-side analysts disseminate their information to other market participant. In doing so, the brokerage firms hope to earn trading commissions. As a result, buy-side analysts face the potential conflicts of working for investment banking firms and the need to generate commissions. In addition to sell-side analysts, there are also buy-side analysts and independent analysts who prepare earnings forecasts. Independent analysts provide their research to a select group of individuals on a contract basis. Buy-side analysts typically work for mutual funds or pension funds or other non-brokerage firms and provide research exclusively for those firms (Gell, 2011, p. 10 f.). This raises the question of whether institutional investors have an information advantage over other investors. According to Groysberg et al. (2008), there is no such advantage. Groysberg et al. (2008) find that the forecasts of buy-side analysts are in fact more optimistic and less accurate than those of sell-side analysts. They attribute this to the higher retention rate for low-quality analysts and the fact that buy-side firms do not measure the performance of their analysts against each other and sell-side analysts (Groysberg et al., 2008, p. 37 f.). They further mention that buy-side analysts are less able to communicate directly with firm representatives (Groysberg et al., 2008, p. 26).

To forecast earnings, financial analysts build financial models that estimate prospective revenues and costs of firms. The model evaluates information about the general economy, the industry and the specific firm and then generates an estimate of the firm's earnings. The weighting of the three sources of information, however, differs between analysts. If the financial analyst believes that the firm is not able to accurately forecast earnings, he is more likely to rely on industry and economic data. (Jennings, 1985, p. 2).

When talking about earnings forecasts, what is meant is usually the consensus earnings estimate. The consensus earnings estimate refers to the mean or median of the forecasts of a group of financial analysts. Typically, financial analysts estimate a firm's quarterly or annual earnings per share (EPS). The more financial analysts provide a forecast estimate, the more accurate the consensus estimate becomes, as extreme and uninformed estimates carry less weight (Barron et al. (1998) in Byard et al. (2011, p. 94)). The accuracy of the forecast also increases with the amount of information available to analysts, their forecasting experience, and their reputation (Klettke, 2013, p. 2). At the beginning of the period, analysts have a higher forecast error compared to right before the earnings release (Capstaff et al., 1995, p. 74). Part of this change in forecast error is due to managers influencing analysts' forecasts by providing them with additional information. (Chopra, 1998, p. 36). Managers have an incentive to revise earnings estimates downward, because missing the consensus earning estimate is associated with a significant drop in stock price (J. R. Graham et al., 2005, p. 3f). Thus, the consensus earnings estimate varies over the year.

Forecast Biases

Financial analysts are consciously or unconsciously subject to biases when making their earnings forecasts. The two most prominent forecast biases in the literature are optimism and herding. Optimism describes the persistent tendency of

financial analysts to issue overly positive earnings forecasts. It is measured as the difference between the consensus earnings forecast and the later realized earnings (Beckers et al., 2004, p. 75). The tendency for optimism has been documented as early as the 1970s and persists until today (McDonald 1973, p. 509; Barefield and Comiskey 1975, p. 244). Dreman and Berry (1995, p. 39) find that the optimism bias is persistent across industries and economic cycles. Financial analysts have incentives to issue more optimistic forecasts. Hong and Kubik (2003, p. 345 f.) note that financial analysts who are more optimistic than the consensus are more likely to experience positive career developments. The reason for this is that investment banks and brokerage houses want analysts to promote stocks in order to generate underwriting business and trading commissions. Athanassakos and Kalimipalli (2003, p. 59) further point out that forecast optimism is the largest at the beginning of the year. As more information becomes available during the year, financial analysts cannot afford to continue being overly optimistic without damaging their reputation.

Herding, on the other hand, describes the social phenomenon of financial analysts to conform and therefore not to deviate too much from the consensus. Scharfstein and Stein (1990) find that herding toward the consensus is less likely caused by fundamental information, but rather a lack of information. Financial analysts who have little or no information tend to herd more (Welch, 2000, p. 371). In addition, the reluctance to deviate from the consensus has been shown to increases with the number of estimates that are close to the consensus and the inaccuracy of analysts' previous estimates (J. R. Graham (1999) and Stickel (1990) in Beckers et al. (2004, p. 75)). This might be explained by the fact that investors view agreement with the consensus as an indication of forecast reliability (De Bondt & Forbes, 1999, p. 144 f.). By simply endorsing the consensus opinion, financial analysts take less reputational risk. As Keynes said, "worldly wisdom teaches that it is better for reputation to fail conventionally than to succeed unconventionally“ (Keynes, 2018, p. 138). De Bondt and Forbes (1999, p. 146) further mentions that herding intensifies with the difficulty and ambiguity of the task. Herding may also be explained by career concerns. Hong et al. (2000, p. 123) find that older analysts are more likely to produce forecasts that deviate from the consensus, while younger analysts tend to be less bold. Similar, Zwiebel (1995, p 2 f.) note that younger financial analysts are more likely to function as opinion leaders due to lower reputational risks. Still, they also note that the earnings revisions from older financial analysts receive more weight due to higher reputational capital.

In the literature there are different explanations for forecast biases. Gell (2011) distinguishes six categories of explanations for biases: cognitive bias, strategic bias, selection bias, news bias, skewed earnings distribution bias and management bias explanation. First, under the cognitive bias explanation, financial analysts are supposed to be irrational and to systematically make mistakes when processing publicly available earnings information. Second, the strategic bias explanation assumes that financial analysts are rational, but produce biased forecasts due to strategic incentives. For example, financial analysts publish positive earnings reports to please a firm's management in order to maintain a good relationship with the firm. Third, the selection bias explanation states that financial analysts make optimistic forecasts only for those firms about which they are truly optimistic because these firms are more likely to bring in trading commissions. For firms that underperform, analysts stop making forecasts, resulting in outdated and hence biased forecasts. Fourth, the news bias explanation attributes forecast optimism to the asymmetric timeliness or earnings due to accounting conservatism. Good news are simply reflected in forecasts in a more timely manner than bad news. Fifth, the skewed earnings distribution bias explanation assumes that financial analysts are truthful, unselective, and rational. Though, they can choose whether to forecast the mean or median of an earnings distribution and thus bias earnings forecasts. Last, the management bias attributes forecast bias to the management practices of accounting discretion and guiding analysts' expectations (Gell, 2011, p. 13 f.).

Forecast Dispersion

Analyst forecast dispersion measures the variation in analysts' earnings forecast for a certain firm and period. Dispersion thus reflects the divergence in analysts' opinion about a firm's future earnings (Han & Manry, 2000, p. 99). Theoretical research shows that forecast dispersion may reflect both uncertainty and information asymmetry (Barry and Jennings 1992, p. 175 f. Barron et al. 1998, p. 422).

Uncertainty arises because financial analysts do not have the exact earnings numbers and instead need to estimate earnings. When earnings are announced, uncertainty decreases. As a result, the dispersion of analysts' earnings estimates typically decreases as well (Barron et al., 2010, p. 332). Likewise, Imhoff and Lobo (1992, p. 437) interpret forecast dispersion as a proxy for ex ante earnings uncertainty. Chopra (1998, p. 38) finds that the dispersion of earnings estimates declines over the year. He attributes the decline in dispersion to quarterly earnings releases and resulting improved visibility of the firm's prospects. Further, Ackert (1997, p. 264 f.) notes that financial analysts issue more optimistic forecasts when uncertainty around the firm is high. However, if the uncertainty is low, financial analysts may hesitate to issue optimistic forecasts due to reputational concerns. For the same reason, financial analysts may also avoid issuing contrarian forecasts when uncertainty is low.

Information asymmetry refers to the differences in information available to financial analysts. In this context, information can be divided into public and private information. On the one hand, public information comprises all firm-related information that is freely accessible to all financial analysts. Private information, on the other hand, is only available to the individual analyst. From the perspective of information asymmetry, forecast dispersion results from the different level of information available to financial analysts

(Barron et al., 2010, p. 332). Ajinkya et al. (1991, p. 393) argue that analyst forecast dispersion is in part due to differential information available to financial analysts at different times. Not all financial analysts prepare and submit their EPS updates at exactly the same time, so there may be a time lag between EPS estimates.

The relationship between public disclosure and forecast dispersion is thus not so obvious. The effect of disclosure depends on whether variations in forecasts are due to differences in information or differences in forecasting models. If analysts share a common forecasting model and observe the same firm-provided information but have different private information, they will attach less weight to their private information as the informativeness of the firm-provided information increases, thereby reducing forecast dispersion. Contrary, if analysts have the same firm-provided and private information, but assign different weights to the constituents of the firm-provided information when forecasting earnings, additional disclosure may increase financial analysts' forecast dispersion. Hence an observed positive association between earnings disclosure and forecast dispersion implies that financial analysts differ in their forecasting models, so that they draw different conclusions from the same observed disclosures. With more disclosures, their earnings forecasts become more dispersed. By contrast, an observed negative relationship between earnings disclosure and forecast dispersion implies that financial analysts vary primarily in their private information (Lang & Lundholm, 1996, p. 471 f.). In addition to the disclosure itself, its quality is also important. Earlier research has shown that poor quality disclosure of financial information is associated with high analyst forecast dispersion. Dechow et al. (1996, p. 27) find that forecast dispersion increases after the disclosure of alleged earnings manipulations. Swaminathan (1991, p. 40) shows that forecast dispersion decreased following the release of Securities and Exchange Commission (SEC)-mandated segment data.

Moreover, (Barron et al., 2010, p. 331) find that levels and changes in forecast dispersion reflect uncertainty and information asymmetry to varying degrees. According to them, levels of forecast dispersion before earnings announcements mainly reflects the variation in uncertainty and not in information asymmetry. Reciprocally, changes in dispersion around earnings announcements reflect variation in information asymmetry rather than variation in uncertainty. This means that when looking at levels of forecast dispersion, i.e. EPS estimates by analysts prior to earnings announcements, it is primarily uncertainty that is responsible for forecast dispersion. But, there is also research that disagrees with the proposition that forecast dispersion reflects uncertainty. Imhoff and Lobo (1992, p. 437) study the dispersion of analysts' forecasts prior to earnings announcements and suggest that the increased forecast dispersion is due to noise in financial statements rather than to uncertainty.

Last, firms with high forecast dispersion experience certain real effects. Han and Manry (2000, p. 119-121) find that firms with high forecast dispersion face high costs of capital and low earnings persistence. Also, Diether et al. (2002, p. 2135-2137) and Johnson (2004, p. 1975 f.) demonstrate that investors pay a premium for stocks with a high dispersion of analysts' forecasts, which leads to lower future stock returns, i.e., the degree of dispersion is negatively associated with future stock returns. Diether et al. (2002, p. 2137-2139) explain this negative association with market friction. In particular, higher dispersion induces a stronger optimistic bias in stock prices, as optimistic investors drive up prices, while pessimistic views are not reflected in stock prices due to short-selling restrictions, causing stocks with high dispersion to be overvalued.

2.3. Limits of Market Forecasting

Financial analysts analyze present firm information and make estimates about the future. These future forecasts are then used by market participants to outperform the overall stock market. It is therefore assumed that the activities of financial analysts add value to financial markets. Economic theory though expresses a great deal of skepticism about financial analysts' ability to forecast market developments. In 1933, Alfred Cowles empirically tested the attempt to predict the development of stock prices. After analyzing 7500 stock market forecasts from financial service providers, he concluded that "statistical tests of the best individual records failed to demonstrate that they exhibited skill, and indicated that they more probably were results of chance" (Cowles, 1933, p. 323). To test whether his results were due to a lack of skill, he repeated his test with the then editor of the Wall Street Journal. Cowles came to the same conclusion. Of 90 forecasts, half were successful, and half were not (Cowles, 1933, p. 323). Kendall and Hill (1953, p. 11) later validated Cowles' (1933) findings by showing that stock prices move randomly rather than predictably. According to economic theory, financial analysts should hence not be able to predict market movements.

The most well-known economic theory is Eugene Fama's efficient market hypothesis. The efficient market hypothesis states that all information that is publicly available about a firm is instantly reflected in a firm's stock price (Malkiel & Fama, 1970, p. 383), which makes long-term prediction of stock market movements impossible. According to Fama, "The Evidence in support of the efficient market model is extensive, and [...] contradictory evidence is sparse" (Malkiel & Fama, 1970, p. 416). For the work of financial analysts, this means that there is not a chance of systematically identifying unpriced information that will be reflected in the stock price at some point in the future. Otherwise, according to the logic of efficient markets, the share price would have already risen (Leins, 2018, p. 21). According to Fama, financial analysts can only predict stock price movement in an efficient market if they have access to insider information not accessible to the general public. Since financial analysts do not regularly possess insider information, the scope to predict market movements appears to be limited (Malkiel & Fama, 1970, p. 413).

Still, research suggests that the market is not always efficient. For example, Jones and Litzenberger (1970, p. 147

f.) find that prices do not respond immediately to the content of quarterly reports. In addition, McKibben (1972, p. 379) shows that publicly available information on sales returns, earnings changes, growth relative to price-earnings ratios, and payout ratios can be used to assemble a portfolio that produces superior returns. Moreover, Jennings (1985, p. 2 f.) argues that if the market were efficient, managerial EPS forecasts should present little new information to financial analysts. In reality, however, EPS estimates differ from those of managers, suggesting that they may not have all the information available. Therefore, the market is not always efficient. The practices of financial analysts thus add value.

2.4. Role of Analysts in Financial Markets

Financial analysts play an important role in financial markets. First of all, financial analysts function as information intermediaries (Beunza & Garud, 2007, p. 15). Market participants have limited attention and resources to analyze firm disclosures (Hirshleifer et al., 2009, p. 2323). By collecting, processing and evaluating information, financial analysts filter out the relevant information. The information is then disseminated either through research reports, recommendations, or earnings estimates. As a result, financial analysts reduce the time and resources market participants need to gather and analyze firm information before making an investment decision. Thus, reducing their transaction costs (Leins, 2018, p. 27). Further, financial analysts help reduce information asymmetry between disclosing firms and market participants (Frankel & Li, 2004, p. 256) and improve market efficiency by incorporating firm-specific information into share prices (Healy & Palepu, 2001, p. 417).

Yet, there are also critical voices in the literature that do not support the proposition that financial analysts are information intermediaries. The view of financial analysts as information intermediaries strongly contradicts with the efficient market theory of Eugene Fama. According to the efficient market theory all publicly available information are immediately reflected in the stock price (Malkiel & Fama, 1970, p. 413-416). Yet, Fama neglects that someone has to first incorporate new information into the market before it can be reflected in the stock price. Financial analysts integrate new information into the market and make sure that the market is efficient in the first place. By collecting, evaluating and distributing firm-related information they become enactors of market efficiency (Leins, 2018, p. 156). Moreover, Hou et al. (2020, p. 3) argues that, in general, financial analysts act as information intermediaries. But that they fail to act effectively as information intermediaries at times because they tend to be overoptimistic, underreact to negative news and overreact to positive news. Higgins and Saito (2007, p. 6) find little support for financial analysts acting as information intermediaries for intangible firms.

Second, financial analysts increase information quality by creating an external layer of scrutiny for financial reporting processes. They monitor firms on a regular basis and scrutinize management behavior and financial reporting irregularities. Thus, further reducing information asymmetry (Yu, 2008, p. 247). Healy and Palepu (2001, p. 408) argue that financial analysts as information intermediaries help to uncover managerial misconduct by engaging in private information production. Yu (2008, p. 268) finds that higher analyst coverage decreases the risk of earnings management. Dyck et al. (2006, p. 2214) suggest that financial analysts are the single most powerful third-party for detecting corporate fraud. Having said that, financial analysts are often under pressure from their employers to secure investment banking business, from competitors to maintain good relationships with managers in order to gain access to private information, and from major clients of their brokerage houses. All of which limits their ability to scrutinize bad firm behavior (Yu, 2008, p. 248).

Third, financial analysts play an important role in the allocation of capital (Wansleben, 2012, p. 421). Although usually portrayed as impartial observers and interpreters of the market, financial analysts actively contribute to the promotion of investments (Leins, 2018, p. 13). Their earnings estimates and investment recommendations generate, increase, reduce, or interrupt the flow of capital. If financial analysts are positive about a firm's future economic outlook, the firm can obtain additional financing on the capital market. However, if financial analysts are pessimistic about a firm's future, their evaluations deprive the firm of capital (Leins, 2018, p. 2 f.). Belnap (2022, p. 6 f.) mentions that reducing processing costs through information intermediaries affects price informativeness, price responsiveness, liquidity, volatility and trading volume.

Fourth, financial analysts perform a role as economic narrators. Through their market forecasts, they give meaning to economic activities in the market and provide a sense of agency for other market participants. This sense of agency allows other market participants to view market activities as predictable, that can be understood through the work of financial analysts, rather than as random. In this way, financial analysts influence the investment decisions of other market participants. Investors become active traders instead of investing passively because they feel they know how the market will develop. Although the market itself is unpredictable (Leins, 2018, p. 157-160).

The next chapter introduces the reader to the concept of ESG investing. The aim of this chapter is to show the reader the value proposition of ESG information and how market participants take it into account when making investment decisions. It theoretically addresses why ESG information matter to market participants. Or, in other words, why there ought to be a association between ESG information and analysts' estimates in the first place.

3. ESG Investing

3.1. ESG Integration in Investment Decision Making

ESG investing refers to the process of considering ESG factors when making investment decisions. ESG investments can be categorized within a broader spectrum of social and

financial investing. On the one side, there is conventional financial investing which focuses on maximizing shareholder value through risk-adjusted financial returns. This investment approach assumes that the efficiency of capital markets will effectively allocate resources to those parts of the economy that maximize benefits, thus contributing to economic development. On the other side, there is pure social investing, such as philanthropy, which only aims at social returns, such as addressing social problems or protecting the environment. Within this spectrum, ESG investing focuses on maximizing financial returns by incorporating ESG factors to help assess long-term risks and opportunities (Boffo & Patalano, 2020, p. 14).

In the last ten years, there has been a growing demand for ESG investing. Bloomberg Intelligence estimates that total global ESG assets may surpass $ 41 trillion in 2022 and $ 50 trillion in 2025, equivalent to one-third of total global assets under management (Bloomberg, 2022). Indicative of this growth are also the more than 4900 signatories to the United Nations Principles for Responsible Investment, with over $ 121 trillion in collective assets under management (Principles for Responsible Investment (PRI), 2022, p. 35). According to one survey, this demand has been driven primarily due to end investors' desire to improve firms' alignment with social and moral considerations. Just about 20% of end investors pursue ESG investing for financial gain or the mitigation of investment risk (Boffo & Patalano, 2020, p. 17). However, institutional investors and asset managers integrate ESG in their investment decision making process primarily to improve long-term risk-adjusted returns and reputation. (BNP Paribas, 2019, p. 13)

ESG criteria can be integrated into the investment process in various ways, with the complexity and the level of integration increasing with each step. The first step, exclusion, simply excluded or avoided firms if their behavior is not aligned with fundamental societal values. Reasons for exclusion might include the manufacturing of controversial weapons or activities that are not aligned with ethical standards, such as tobacco, alcohol or gambling. The second step, standards-based or inclusive screening, aims to include or give greater representation to firms that meet international standards such as the UN Sustainable Development Goals. For instance, a firm may be included based on its ESG performance relative to its peers (best-in-class) or because it exceeds a certain ESG score. The third step is similar to the inclusive screening, in which ESG ratings are used to rebalance portfolio exposure to firms with higher ESG ratings and away from firms with lower ESG ratings. The fourth step is to focus on a particular E/S/G pillar and the underlying metrics. For example, a fund may focus on environmental issues and in particular on the carbon footprint and intensity of firms. The final step is full ESG integration, which means the systematic and explicit inclusion of ESG risks and opportunities in the investment process. ESG factors are continuously considered throughout the investment process (Boffo & Patalano, 2020, p. 32 f.)

With respect to ESG investing, one common question that arises, is whether there is a trade-off between ESG investing and traditional investing in terms of returns. On the one side, Berk and van Binsbergen (2021, p. 2) suggest that investors have non-financial preferences for green stocks and are therefore willing to accept lower returns for owning green stocks. Cornell (2020, p. 7) argues that investors buy green stocks as a hedge against ESG-related risks and are willing to accept lower expected returns in return. According to Cornell (2020, p. 6 f.), investors can only profit from green stocks if they are undervalued due to positive undisclosed ESG information. Raghunandan and Rajgopal (2022, p. 35) show that ESG funds appear to underperform financially relative to other funds within the same asset manager and year. On the other side, Kempf and Osthoff (2007, p. 13 f.) find that an investment strategy based on buying stocks with high socially responsible ratings and selling stocks with low socially responsible ratings leads to abnormal returns. Nagy et al. (2016, p. 121) find that portfolios that incorporate ESG into their decision making outperform the MSCI World Index over the sample period. Then again, Fish et al. (2019, p. 13) show that little difference existed between the returns of ESG-weighted and non-ESG-weighted portfolios. In addition, JP Morgan (2016) notes that the yearly net returns of the MSCI World Benchmark Index and the MSCI World ESG are not much different (JP Morgan (2016) in Boffo and Patalano (2020, p. 36)). Therefore, findings on ESG investment performance in the last 15 years are mixed.

3.2. Demand for ESG Information

The growth in ESG investing is accompanied by an increased demand for ESG information (see Figure 3 in the annex), research and ratings in order to make informed and meaningful investing decisions. Whether it be assessing a firm's economic long-term position or its impact on society. ESG refers to the three non-financial pillars that firms are expected to report in. The goal of ESG is to capture all non-financial risks and opportunities associated with a firm's daily operations. As mentioned earlier, investors increasingly demand ESG information to assess a firm's social impact or long-term risk-adjusted returns. But, it is not always clear what falls under these pillars. This is because there is no standard ESG reporting framework yet. For this reason, firms are typically applying one or more frameworks to determine how and what they want to report on. The most commonly used frameworks are the Global Reporting Initiative (GRI) and the Sustainable Accounting Standards Boards (SASB)' standards (Deloitte, 2022). The environmental pillar can include issues such as natural resource use, carbon emissions, energy efficiency, pollution, and sustainability initiatives. (Boffo & Patalano, 2020, p. 21). In terms of resource use, for example, a firm could report whether it uses new or recycled materials in its production and if it ensures that its products are recycled or end up in a landfill. A firm could also report on land use practices, such as deforestation and biodiversity disclosure, or water use. The social pillar can include issues regarding workforce-related practices, human rights,

diversity and supply chain. For instance, a firm could report on how it manages their employee development and labor practices. They could also report on product liability related to the safety and quality of their products. Or a firm could report on supply chain labor, health and safety standards or controversial sourcing issues. The governance pillar can include issues such as board independence, board diversity, shareholder rights, management compensation and corporate ethics. For example, a firm could indicate whether management compensation is linked to the firm's sustainability performance or whether it has implemented measures to prevent anti-competitive practices and corruption (Deloitte 2022; Boffo and Patalano 2020, p. 21).

In a 2020 survey conducted by the SustainAbility Institute, investors were asked about their most important sources of ESG information. Investors indicated that corporate ESG ratings (55%), direct contact with firms (55%), corporate sustainability reports (50%) and internal research (41%) were the most useful sources of ESG information. Of the investors surveyed, 96% said they use ESG ratings, with 65% using them at least once a week (The SustainAbility Institute, 2020, p. 17 f.). Another survey conducted by Ninety One finds that 88% of professional fund managers currently use ESG ratings, with 92% expecting to increase their use (Ninety One, 2022). ESG rating agencies therefore play an important role in the ESG investment ecosystem, as investors rely on their assessment of ESG information.

The next chapter introduces the reader to ESG rating agencies. The chapter aims to inform the reader about the role ESG rating agencies play in the financial markets and to give him an overview of the ESG rating industry. In addition, this chapter intends to briefly introduce the three most important ESG rating agencies in this master thesis.

4. ESG Rating Agencies

4.1. Objectives and M&A activities

ESG rating agencies are third-party information intermediaries that assess a firm's ESG risks and opportunities based on public information and sometimes private surveys (Scalet & Kelly, 2010, p. 71). Due to the complexity and amount of information available, investors and other stakeholders rely on ESG rating providers for their assessments (The SustainAbility Institute, 2020, p. 6). The rating process is not always transparent, although ESG rating agencies attempt to be more transparent by disclosing their rating methodologies (Boffo & Patalano, 2020, p. 64 f.). The result is an ESG rating, often accompanied by a research report that provides additional information about the analysis. The ESG rating informs stakeholders about a firm's ESG performance. Investors usually pay to get access to these ratings. ESG ratings can be quit expensive, so that access tends to be limited to a select group of professional investors (The SustainAbility Institute, 2022, p. 10). Despite that, there are a number of ESG ratings that can be accessed online for free.

It is often misunderstood what ESG rating agencies actually measure. ESG rating agencies provide insights into a firm's ESG quality. Unfortunately, there is no single agreed upon definition of what is considered ESG quality. A common misperception is that ESG reflects the impact a firm has on the wellfare of its stakeholders, such as employees, suppliers, customers, local communities, and the environment. According to this view, firms can enhance their ESG rating by discontinuing activities that are harmful to stakeholders or by improving their business practices in affected areas for the benefit of stakeholders. This view therefore assumes that ESG quality measures the impact the firm has on societal and environmental factors In reality, however, the opposite is true. ESG quality measures the influence that social and environmental factors have on the firm and whether these factors are financially material. Through strategic planning, targeted investments or changes in operations, the firm can address these risks and opportunities. In the short term, this will lead to higher costs, but in the long term it will strengthen the firm's financial position (Larcker et al., 2022, p. 2).

ESG rating agencies have stated objectives. A common stated objective of ESG rating agencies is the reduction of investment risk. This objective assumes that ESG quality increases financial performance by reducing ESG factors that pose a risk to the firm's business model or operations. For this purpose, MSCI argues that its ratings "support ESG risk mitigation and long-term value creation". Likewise, Sustainalytics states that it measures "the degree to which a firm's economic value is at risk" due to ESG factors. If this is true, it would mean that firms with high ESG quality would face fewer regulatory violations, litigation, or bankruptcies in the future. Another stated objective of ESG rating agencies is that their ratings can predict returns. HIP contends that its ESG ratings "correlate with better returns for the same amount of risk". Arabesque asserts that their approach "is all about identifying firms that are better positioned to outperform over the long term" and that their algorithm for ESG ratings "will only use information that significantly helps explain risk adjusted performance". If this is true, an ESG rating upgrade should be associated with a subsequent change in stock price. Other stated objectives include measuring a firm's societal impact (ISS) and transparency and commitment to ESG (Refinitiv) (Larcker et al., 2022, p. 3).

In the last 15 years, there have been a large number of mergers and acquisitions (M&A) in the ESG rating market. As a result, the ESG ratings market has become increasingly consolidated. There are two main reasons for this surge in M&A activity. First, as the ESG ratings market matured, established rating providers entered the field and began acquiring smaller ESG firms to obtain expertise and market share. Recent examples include Moody's acquisition of a majority stake in Vigeo Eiris, S&P Global's purchase of RobecoSAM or Fitch's development of its sustainability platform. Second, increasing investor demand for broader and deeper information and the complexity of ESG reporting forced established ESG rating providers to expand their product offerings to remain competitive. Therefore, existing ESG ratings agencies merged or acquired smaller ESG firms (The SustainAbility Institute, 2020, p. 6). Currently the ESG rating

market remains highly competitive, with the quality, range, scope, and frequency of ESG ratings increasing to the benefit of investors (The SustainAbility Institute, 2022, p. 3). Some of the most important ESG ratings agencies at present include Bloomberg, CDP, FTSE Russell's ESG Ratings, ISS, MSCI, Sustainalytics, Bloomberg, Refinitiv (formerly Thomson Reuters), RobecoSAM, which are used by investors primarily for their broad coverage of firms (The SustainAbility Institute, 2020, p. 33-35).

4.1.1. Sustainalytics

Sustainalytics is a leading independent ESG rating agency that evaluates and rates more than 20.000 firms worldwide based on their ESG performance. Sustainalytics is part of the Morningstar Group, which acquired the firm in 2020 (Cision, 2020). At present, Sustainalytics employs more than 1800 staff in 17 offices worldwide, including more than 800 research analysts with varied multidisciplinary expertise (Sustainalytics, 2022).

Sustainalytics has completed several key mergers and acquisitions over the past years. In 2008, the firm was formed from the consolidation of Analistas Internacionales en Sostenibilidad (Spain), Dutch Sustainability Research (Netherlands) and Scoris (Germany). In 2009, Sustainalytics merged with Jantzi Research, whose CEO Michael Jantzi serves as the current CEO (Novethic, 2014, p. 21). In 2012, Sustainalytics acquired Responsible Research, a ESG research firm based in Singapore and Share Dimension. In 2018, Sustainalytics acquired Solaron, another provider of ESG research and ratings, and in 2019 GES, a specialist in engagement, screening and fiduciary voting services for institutional investors. Last, in 2020, Sustainability acquired OMX, a supply chain data platform that tracks the socioeconomic impact of supply chains. As a result of these acquisitions, Sustainability is further strengthening its market position as a sustainability service provider (Sustainalytics, 2022).

Sustainalytics offers investors a wide range of products and services to help them navigate ESG-related risks and opportunities. Among others, Sustainalytics offers ESG Risk Ratings, Carbon Risk Ratings, Product involvement data, Controversy Research, Global Standards Screening data and an Impact Metrics (see Table 13 in the annex).

4.1.2. MSCI ESG Research

MSCI ESG Research is a provider of in-depth research, ratings, and analysis of ESG business practices of more than 10.000 firms. MSCI ESG Research is a fully-owned subsidiary of MSCI, a provider specializing in tools and services for investment decision support. The firm currently employs 600 ESG employees, including its foreign affiliates, with approximately 250 analysts and researchers worldwide (MSCI ESG Research, 2022, p. 3).

MSCI ESG Research was formed through the acquisition of RiskMetrics by MSCI in 2010. RiskMetrics has itself acquired the governance consulting agency ISS (US) in 2007 and the two rating agencies Innovest (US) and KLD (US) in 2009. In addition, MSCI acquired the governance service agency GMI ratings in 2014 and the climate change scenario analysis firm Carbon Delta in 2015 (MSCI 2019, p. 1; Novethic 2014, p. 14) . As a result, MSCI has further strengthened its focus on ESG analysis.

MSCI Research offers a wide range of ESG-related products and services to investors. These include the following products and services: MSCI ESG Ratings, MSCI ESG Controversies, MSCI ESG Global Norms Screening, MSCI Climate Value-at-Risk, MSCI ESG Business Involvement Screening Research, MSCI ESG Portfolio Analysis (see Table 14 in the annex; MSCI ESG Research 2022, p. 3 f.). MSCI has further set up a custom division to handle special client requests (Novethic, 2014, p. 14).

4.1.3. Refinitiv (formerly Thomson Reuters)

Refinitiv is a global provider of financial market data and infrastructure and offers ESG scores for more than 11800 firms worldwide. Refinitiv is a subsidiary of the London Stock Exchange Group (Refinitiv, 2022a). The firm currently employs over 350 content research analysts trained to collect ESG data (Refinitiv, 2022c).

Refinitiv in its current form is the result of several key mergers and acquisitions. In 2009, Thomson Reuters acquired Asset4, a Swiss firm that provides a global database of ESG information. Asset4 provides research on financial and non-financial information and was the first firm to provide raw ESG data that could by used by investors. In 2010, Thomson Reuters also acquired Point Carbon, an information provider specializing in energy and the carbon market (Novethic, 2014, p. 5 and 27). In 2018, Refinitiv was created through the acquisition of Thomson Reuters Group's Financial & Risk Division by investment firm Blackstone. Blackstone acquired a 55% stake in the newly formed firm, while Thomson Reuters retained a 45% stake (Reuters, 2018). In 2019, the London Stock Exchange Group acquired Refinitiv (The Economist, 2019).

The following chapter introduces the reader to the methodology of ESG rating agencies. The reader learns about the process of compiling ESG ratings. The purpose of this chapter is to make the reader aware of where ESG rating divergences may occur.

4.2. Methodologies of ESG Rating Agencies

4.2.1. Information Input and Firm Disclosure

ESG rating agencies collect a broad range of public and non-public information about the firm and its industry to assess a firm's ESG performance. The kind of information that is collected is important us, as it reveals what is factored into ESG ratings. As Dillenburg et al. (2003, p. 170) emphasize "what gets measured, gets managed".

In the case of ESG rating agencies, ESG data is collected in various ways, such as through corporate social responsibility reports, voluntary firm surveys and questionnaires, analysis of media reports, independent research, and active communication with the management of the rated firm and stakeholders (i.e. non-governmental organizations (NGOs), trade

unions, governmental organizations, etc.) (Scalet & Kelly, 2010, p. 71). However, ESG rating agencies differ in what counts as relevant data, which makes it difficult to compare the results of different rating agencies. This is also the case with Sustainalystics, MSCI and Refinitiv. Sustainalytics collects data from firm's public disclosure, the media, and NGOs (Sustainalytics, 2020, p. 7). MSCI gathers macro data, firm disclosures and data from media, NGOs, and other stakeholders (MSCI, 2022b, p. 14). Refinitiv fully automatically collects publicly available data from annual reports, firm websites, NGO websites, stock market reports, CSR reports and news sources (Refinitiv, 2022b, p. 4). In comparison to financial disclosures, ESG data is largely unstandardized, frequently unstructured, difficult to compare, and tends to be more subjective than financial disclosures (Sipiczki, 2022, p. 6).

None of the investigated rating providers conduct surveys and questionnaires. Examples of ESG rating agencies that conduct subjective surveys or questionnaires are S&P Global, CDP and RobecoSAM (Deloitte, 2021). Surveys have the disadvantage compared to data-driven approaches that they are subjective in nature. What data is collected will depend on who is creating the survey. Also, responsibility for answering the questions truthfully resides with the firm. Self-reported survey data raises doubts regarding reliability (Sipiczki, 2022, p. 6). In addition, non-response rates are high (Chatterji & Levine, 2006, p. 30). Firms are understaffed to deal with the complexity and volume of ESG data requested. Managing a firm's ESG ratings can require hundreds of hours and several dedicated staff, something that even large organizations may struggle to accomplish. Small and mid-sized firms with fewer resources may risk not managing ESG ratings at all (The SustainAbility Institute, 2020, p. 7).

A major challenge for ESG rating agencies is data completeness. ESG rating agencies' models comprise of hundreds of material input variables for which data is required. As mentioned previously, corporate data is an important source of information for ESG rating agencies. In 2011, though, just under 20% of S&P 500 firms reported on their sustainability efforts, corporate social responsibility activities, and ESG performance. At the time corporate data on ESG was scarce. Since then, things have improved significantly. Last year, 96% of S&P 500 firms published a sustainability report. The number of non-reporters thus dropped to 4% (Governance & Accountability Institute, 2022, p. 5).

Another major challenge for ESG rating agencies is data consistency. A report by Deloitte that studied 4000 ESG reports finds a significant number of data omissions, groundless claims and inaccurate figures (Hespenheide & Koehler, 2013, p. 12). Firms may report on different ESG issues, because they consider different issues material to the firm's financial performance. This will lead to missing data in ESG rating agencies' models. Consequently, ESG ratings agencies face the challenge of determining how to address missing data. One approach would be to exclude data points with missing information. But, this would make it difficult to compare the ESG results of firms that report on certain ESG issues with those that do not. Another approach would be to make an assumption about what the data might be. ESG rating agencies fill data gaps by drawing on the opinions of industry peers, making assumptions, or obtaining missing information from third-party sources (Sipiczki, 2022, p. 7). For example, MSCI appears to assume that the firm's performance is in line with the industry average when no information is available. In contrast, FTSE makes the assumption that the firm's performance is the worst in the case of missing data to encourage information transparency. A third approach would be to use statistical methods to impute missing values (Larcker et al., 2022, p. 5). Still, all these approaches only reduce the problems of data inconsistencies, but do not eliminate them. Another issue that causes data inconsistencies is differences in reporting metrics and scales. When firms report on the same information but use different methods, scales or metrics, the information is not directly comparable. For example, one firm may report on workplace safety based on the number of incidents, while another may report the number of injuries over a period of time, and yet another may report how much time was lost due to workplace injuries (Larcker et al., 2022, p. 5). This short example illustrates the importance of standardized measurements. One way to eliminate data inconsistencies is to standardize corporate reporting on ESG issues. Firms and ESG rating agencies can increase data consistency by voluntary adopting or aligning themselves with sustainable reporting frameworks, such as the GRI and the SASB's standards (Boffo & Patalano, 2020, p. 20). Government disclosure mandates such as the EU Non-Financial Reporting Directive (NFRD) (EU-Directive Nr. 2014/95/EU) or the SEC's proposed ESG reporting mandate may also help to increase consistency in ESG reporting (Securities and Exchange Commission (SEC), 2022). Then again, the UN Principles for Responsible Investment (PRI) and International Corporate Governance Network (ICGN) suggest hat given the heterogeneity of users of ESG information, there is no one set of metrics or one framework that could satisfy all users. In their opinion, firms should disclose basic standardized ESG information and complement it with more customized ESG reporting (UN PRI and ICGN, 2018, p. 2).

Recent advances in artificial intelligence (AI) are contributing to changes in ESG rating agencies' data collection processes. ESG rating agencies are now deploying computer algorithms to automate data collection tasks and analyze information (Brackley et al., 2022). Repeatable tasks are carried out by bots in a fraction of the time, and algorithms can read information that otherwise might have been unusable due to its size or amount of low-quality data (S&P Global, 2020). While ESG rating analysts mostly draw on corporate ESG disclosures, algorithms also evaluate a wide range of media news (Nomura, 2022, p. 2). AI also enables ESG rating agencies to analyze data more efficiently. AI can help the agency exclude firm statements that mention ESG practices, which are not material to the business and hence unlikely to matter to investors (S&P Global, 2020). ESG ratings agencies that utilize AI technology can also update their data daily, re-

ducing inaccuracies (Nomura, 2022, p. 2). Natural language processing is a particularly promising area. Sentiment analysis algorithms allow AI to detect the tone of a conversation. For example, a program trained to read transcripts of firm's quarterly earnings calls can use natural language processing to assess the tone of the CEO's words to gauge how engaged a firm appears on ESG issues. Using this approach, ratings agencies can deliver an in-depth overview of a firm's stance on ESG (S&P Global, 2020). In theory, AI gives ESG rating agencies access to a greater amount of higher quality data to incorporate into their ESG ratings. However, there are also challenges related to AI technology. For instance, it is difficult to substitute an ESG analyst who speaks directly to a firm or market participants and then makes a nuanced assessment. (Nomura, 2022). In addition, the opacity of the algorithms makes it difficult to assess the validity of the data. (Lu, 2021, p. 158). AI should therefore be viewed as useful tool to complement traditional data assessments and help to address human biases while im-proving understanding of evaluation results.

4.2.2. Information Processing and Transparency

For the most part, ESG raters are reluctant to reveal the inner workings of their ESG ratings. ESG rating agencies are reluctant to share their methodological process to ensure that firms cannot game the system. Yet, without some methodological openness, investors may lack confidence in the ESG rating process and doubt its usefulness for investment decisions. This is because analysts are unable to assess whether the rating truly reflects the firm's ESG performance (Brackley et al., 2022). Perhaps in response to recent regulatory scrutiny and increasing criticism of their role in the market, ESG rating agencies are moving towards greater methodological transparency (Brackley et al., 2022). While this move is commendable, ESG rating agencies do not currently disclose their methodologies in a fully transparent manner (Doyle, 2018, p. 8). Since the business model of ESG rating agencies is based on product differentiation, it is unlikely that the opacity of ESG rating agencies is going to change entirely unless there is a statutory mandate (Sipiczki, 2022, p. 6).

In the wake of increased transparency, it has become evident that ESG rating agencies' approaches to ESG ratings can vary widely. ESG rating agencies can differ in which input variables are relevant for measuring ESG performance, what constitutes as relevant data for these input variables, and how these input variables are weighted. Due to this differences between the methodologies used by ESG rating agencies, their ESG ratings could tell very different stories about firms‘ ESG performance (Scalet & Kelly, 2010, p. 72). Most often, however, institutional investors are less interested in the ratings itself than in the differences in methodology to form opinions on important issues for their own ESG analyses. The analysis of ESG rating agencies' methodological approaches is thus beneficial as it helps investors gain valuable insights into what factors determine the final ESG ratings (The SustainAbility Institute, 2020, p. 44 f.).

The process of creating an ESG rating usually looks something like this. First, certain key issues are defined and the corresponding data collected. After that, input variables that are not relevant for the business or the industry are removed. Next, the selected key issues are aggregated into subcategories and categories, then into E/S/G pillars, and finally into the rating itself. During this process, various weightings are applied as the discretion of the ESG rating agency (MSCI, 2022b, p. 2).

ESG rating agencies define the type and number of input variables they want to use for their ESG analysis. The number of input variables is usually very large, with hundreds and sometimes thousands of variables (Larcker et al., 2022, p. 4), with a wide variety of definitions due to a lack of regulation (Sipiczki, 2022, p. 6). Several ESG rating agencies use specific ESG framework providers such as GRI, SASB and Task Force on Climate-Related Financial Disclosure (TCFD) to select individual input variables (Boffo & Patalano, 2020, p 31). This way, they can increase the transparency and comparability of their ratings. But, Escrig-Olmedo et al. (2019, p. 14) find that ESG rating agencies are not fully integrating sustainability principles into their rating process. Input variables are also selected to some degree based on data availability to ensure that ESG rating agencies can measure can accurately measure each indicator over time. In addition, input variables can differ significantly across industries to reflect financial materiality or because ESG rating agencies simply consider different aspects of ESG to be financial material (Boffo & Patalano, 2020, p 31).

ESG rating agencies also change input variables in their ESG model over time to assess ESG performance in a more robust and accurate way. Escrig-Olmedo et al. (2019, p. 14) examine how the input variables of ESG rating agencies have changed between 2008 and 2018. They find that ESG rating agencies have mainly integrated new environmental and governance input variables. They further find that ESG rating agencies are now integrating more complex and integrated input variables such as data security and privacy as well as supply chain management (Escrig-Olmedo et al., 2019, p. 11).

The categories and subcategories of ESG rating agencies are quite similar (Boffo & Patalano, 2020, p 31). Occasionally however, there are slight differences between ESG rating agencies. These differences mostly result from different labels of categories and the different assignment of subcategories (see Table 15 and 16 in the annex). In some cases, though, ESG rating agencies also differentiate themselves from one another by including additional variables. MSCI, for example, includes social opportunities and stakeholder opposition in its rating process, while Refinitiv includes data privacy and CSR strategy. But, in most cases areas of interest are overlapping (see Table 15 and 16 in the annex).

To compile the input variables into subcategories, categories, pillars and finally ESG ratings, different weights are applied. One approach would be to simply apply equal weights to all input variables, subcategories, etc. This has the benefit of being simple, transparent and more compara-

ble across industries (Nagy et al., 2020). Still, not all inputs may be equally material to firms' ESG performance. As a result, ESG rating agencies often apply different weights to different variables and categories and sometimes even to different pillars based on financial materiality (Larcker et al., 2022, p. 4). Those weightings are mostly based on subjective judgments, even though various ESG rating agencies rationalize their decisions (Boffo & Patalano, 2020, p 31). Another approach is to apply an optimized weighting based on historical data. In this approach, the weights are adjusted to mirror the best financial performance based on a collection of historical data. For example, a research report from MSCI finds that weights of 25% E pillar, 5% S pillar and 70% G pillar yield the best financial results. Yet another approach is to apply industry-specific weights. The advantage of this approach is that it more precisely reflects industry exposures to E, S, and G risks. But, the disadvantage of this approach is that it leads to more complexity and is less comparable across industries. The same research report from MSCI notes that the E pillar weighting varies from 5.8% for the communications services sector to 62.1% for utilities. The S pillar weighting varied between 16.3% for the energy sector and 59.8% for the financial sector. In the short term, Nagy et al. (2020) find that both equal-weighted and optimized approaches demonstrated superior performance, attributed to increased exposure to governance issues. In the longer term, however, the industry-specific weighted approach showed the strongest financial performance (Nagy et al., 2020). Because ESG measures the long-term risks and opportunities to a firm's financial performance, the industry-specific weighted approach hence appears superior.

ESG rating agencies sometimes also incorporate controversies surrounding rated firms into their ESG ratings. Controversies are events that cause reputational damage and demonstrate a firm's lack of preparedness and/or inability to deal with emerging events and risks. Having said that, not all ESG rating agencies include controversies into their ESG ratings. Some provide controversies as a stand-alone rating that exists alongside the pillars and contributes to a combined overall ESG rating, while others do not consider controversies at all (Boffo & Patalano, 2020, p 30). In summary, there are methodological parameters that allow ESG rating agencies to produce different ESG ratings. In addition, ESG rating agencies are not fully transparent about how their ratings are produced.

4.2.3. ESG Ratings and Rating Biases

A rating is an evaluation provided by a third party. It is an information product, a statement created with the explicit purpose of being communicated outward (Poon, 2012, p. 460). In the case of ESG ratings, ESG rating agencies provide an evaluation of a firm's ESG performance, which they communicate to investors and other stakeholders.

ESG ratings are usually expressed in the form of letters or numbers. Some ESG rating agencies use a seven-point scale from AAA to CCC. Others use a twelve-point scale from A+ to D, similar to grades in the Anglo-American education system. Yet others publish scores on a percentile basis using a scale of 1 to 100, where 100 can either represent high ESG quality (positive) or high ESG risk (negative). In addition, many ESG rating agencies claim to measure industry-relative ESG performance, while some claim to measure absolute ESG performance.

Industry-adjusted ratings enable investors to compare ESG performance among firms operating within the same industry. In this way, firms can be compared against their industry-peers in their ability to manage financial material ESG risks. However, ESG ratings based on industry criteria hinder the ability to compare firms across different industries and are highly dependent on the assigned industry. In contrast, absolute ESG ratings can be compared across industries. Although ratings may vary depending on the industry to which firms are assigned. Firms in more sustainable industries tend to receive higher ratings, while firms in less sustainable industries tend to receive lower ratings (Larcker et al., 2022, p. 3 f).

Moreover, ESG ratings are expensive. Institutional investors spend on average $ 487,000 per year on external ESG ratings, data, and consultants. Many use more than one ESG source in their investment process (The SustainAbility Institute, 2022, p. 5). A 2021 survey finds that more than half of institutional investors use more than one ESG data and research source, with 25% anticipating to use six or more sources in the next two to three years (Capital Group, 2021, p. 29). There has also been discussion about whether ESG performance can be distilled into a single rating. Some investors hide behind ESG ratings and use them as a substitute for in-depth ESG research and analysis. They may see ESG ratings as a quick fix. This happens because some investors may lack the resources for fundamental research or simply want to check a box (The SustainAbility Institute, 2020, p. 31). Other investors stress that ESG performance can not be aggregated into a single rating and that additional in-house research is needed to make sense of ESG ratings (The SustainAbility Institute, 2022, p. 30). These investors view ESG ratings as a starting point to help them understand the broader landscape and to benchmark firms against each other. For instance, a poor rating may signal the need for further research. They rely on their own thinking and use ESG ratings for the underlying data rather than the scores themselves. They develop a strong sense of which ESG factors are the most important for a particular industry and then perform their own evaluation of a firm's ESG performance (The SustainAbility Institute, 2022, p. 23 f.).

ESG ratings are also regularly biased. The most prevalent biases are firm size, geographical bias, and industry affiliation. One pattern is that ESG ratings are biased towards larger-sized firms. Firms with higher market capitalization or free float are more likely to be covered by raters, and their ratings are more likely to be reassessed. Recent initial public offerings are unlikely to be rated in their first year of listing (Brackley et al., 2022). Unlisted firms are often excluded from ESG ratings completely (Zhang, 2021). Larger firms also tend to receive higher average ratings compared

to smaller firms (Giese et al., 2019, p. 77). The reason for this might be that larger firms disclose more ESG data due to more designated employees or the adoption of better sustainable management tools (Drempetic et al., 2020, p. 153). A second pattern is related to the firm's location. Firms listed on exchanges in North America and Europe are far more likely to get properly rated than those trading elsewhere, particularly in emerging markets (Brackley et al., 2022). Moreover, firms in Europe regularly achieve higher ratings on average than firms in the US. This pattern is not due to higher quality ESG practices by European firms, but rather to mandatory reporting requirements. Firms in the EU are required by law to report on various environmental and social topics under the Non-financial Reporting Directive and Corporate Sustainability Reporting Directive. As a result, there is greater availability of non-financial information (LaBella et al., 2019, p. 5). The third pattern is industry-based. ESG ratings that are not industry-adjusted, i.e., do not assign scores based on industry peers, may assign higher average scores to certain industries (such as banking and telecommunications) and lower scores to others (such as tobacco and gambling) (Larcker et al., 2022, p. 5). Furthermore, industry-weighted ESG ratings assume that firms in the same industry have similar business models and are therefore exposed to similar ESG risks and opportunities. However, this approach can cause oversimplification in cases where firms are not comparable. While it is important to standardize methodologies, without individualized weightings, ESG ratings might be skewed (Sipiczki, 2022, p. 8).

In addition, research shows that ESG ratings have moved upward over time. D. E. Shaw (2022) analyze MSCI's aggregate ESG scores for all Russel 1000 firms between 2015 and 2021, and find that scores have improved by 18% over this period. Still, structural changes, such as changes in index composition, changes in component weighting, and greater disclosure by firms, account for only 6% of this improvement. The remaining 12 are not explained by MSCI. D. E. Shaw (2022) attribute this gap to grade inflation (D. E. Shaw, 2022, p. 6). Other research shows that low scoring firms have seen greater improvement in their ESG ratings than high-scoring firms. They attribute this greater score improvement to increased investors scrutiny (Boffo & Patalano, 2020, p. 43).

Furthermore there are a number of issues that affect the quality of ESG ratings. First, there is a conflict of interest arising from the provision of consulting services to rated firms. The practice of offering paid services to rated firms raises serious concerns about the independence of those ESG ratings (Larcker et al., 2022, p. 7). Tang et al. (2022, p. 29) find that firms affiliated with ESG rating agencies receive higher ESG ratings than firms not affiliated with them. Second, ESG ratings are mostly backward-looking, i.e., they evaluate past performance, while investors actually look for indicators of future performance (The SustainAbility Institute, 2020, p. 28). As a result, investors have stated that they would like to have more timely updates (The SustainAbility Institute, 2020, p. 43). Third, recent research claims that ESG ratings do not reliably predict future sustainability performance and do not correlate with ESG risk management capabilities (Brackley et al., 2022). Fourth, investors report that ESG rating agencies often do not respond to complaints about inaccurate information from rated firms. ESG rating agencies are often not sufficiently staffed to provide comprehensive support (The SustainAbility Institute, 2020, p. 28).

4.2.4. Comparison of ESG Rating Agencies' Methodologies

Sustainalytics

In terms of methodology of ESG ratings, Sustainalytics assesses a firm's ESG performance by measuring the extent to which a firm's economic value is exposed to unmanaged material ESG risks (Sustainalytics, 2021, p. 4). The analysis is based on data collected from a firm's public disclosure, the media, and NGO reports. The model includes between 70-90 ESG indicators for large and mid cap firms and between 20-30 for small cap firms. Indicators are selected based on their relevance to the assigned peer group and to the firm's particular business model. At the moment, Sustainalytics distinguishes between 138 peer groups, which are categorized into 42 distinct industries. Sustainalytics uses building blocks that start with corporate governance, consider material ESG issues, and then look for idiosyncratic ESG issues. Betas are then used by embedding the impact of events on financial performance into the process (Sustainalytics, 2021, p. 5-8). Once the analysis is done, firms have two weeks to provide feedback and submit additional information. The final result compiled into a score between 0 and 100, with a lower score being better as it means less exposure to unmanaged ESG risks (Sustainalytics, 2020, p. 7). The rating is absolute, meaning it is comparable across all peer groups covered. (Sustainalytics, 2021, p. 4). In addition, Sustainalytics provides individual E/S/G cluster scores and controversy research. Those are not used to calculate the ESG Risk Rating but provide investors with additional information on ESG performance (Sustainalytics, 2021, p. 12-14). The ratings are updated annually, while controversy research is updated as events occur (Sustainalytics, 2020, p. 5).

MSCI ESG Research

MSCI's rating methodology is as follows. First, MSCI collects macro data, firm disclosures and data from media, NGOs, and other stakeholders (MSCI, 2022b, p. 14). Then, MSCI measures a firm's exposure to material ESG risks and the quality of a firm's risk management (MSCI, 2022b, p. 6). This is done by analyzing the individual E/S/G pillars based on a selection of 35 key issues. Firm-specific exceptions are allowed for firms with diversified business model, facing controversies, or based on industry rules (MSCI, 2022b, p. 3 f.). Figure 4 in the annex shows an example of chosen key measures for the Coca Cola. Each environmental and social key issue typically accounts for 5% to 30% of the total ESG rating. The weightings take into account the industry's contri-

bution, relative to all other industries, to negative or positive environmental or social impacts, as well as the timeframe in which the risk or opportunity is expected to materialize. The weight of the governance pillar is set at a minimum value of 33% (MSCI, 2022b, p. 5 f.). Controversies are directly included in the rating to indicate structural problems in a firm's risk management (MSCI, 2022b, p. 9). MSCI is proactively reaching out to firms for feedback. But, they do not issue surveys or questionnaires or conduct general interviews with firms. Neither are information that is not publicly available to stakeholders accepted and taken into account (MSCI, 2022b, p. 14).To arrive at the final ESG rating, the weighted average of the E/S/G pillar is computed and then normalized relative to industry peers. The best possible score is AAA and the worst CCC. The rating is intended to be interpreted relative to a firm's peers and not absolute (MSCI, 2022b, p. 10 f.). After the rating is published, firms are monitored on a systematic and ongoing basis. Controversies are monitored on a daily basis and new information is reflected in reports on a weekly basis. Significant changes to scores trigger a review and rerating (MSCI, 2022b, p. 14).

Refinitiv

Refinitiv ESG scores are designed to transparently and objectively measure a firm's relative ESG performance, commitment and effectiveness (Refinitiv, 2022b, p. 3). The methodology is as follows. Refinitiv's model is fully automated, data-driven, and transparent, making it free from subjectivity and hidden calculations and inputs (Refinitiv, 2022b, p. 6). The analysis is based exclusively on publicly available data from annual reports, firm websites, NGO websites, stock exchange filings, CSR reports and news sources (Refinitiv, 2022b, p. 4). The model captures and calculates over 630 firm level ESG measures, of which a subset of 186 of the most comparable and material are used for the firm valuation and scoring process (Refinitiv, 2022b, p. 6). Indicators that are irrelevant for a particular sector are excluded (Refinitiv, 2022b, p. 9). Not reporting on immaterial data points has no significant influence on a firm's rating, however, not reporting on highly material data points has a negative impact on a firm's rating (Refinitiv, 2022b, p. 3). The ESG measures are then aggregated into categories. Environmental and social categories are benchmarked against other firms in the same industry, whereas governance categories are benchmarked against other firms in the same country of incorporation. Categories are then compiled into weighted E/S/G pillars from which the final ESG score is calculated (Refinitiv, 2022b, p. 8 f.). Investigated firms are not asked for feedback, although they may request updates at any time (Deloitte, 2021). Refinitiv has two different scores. The regular ESG score and the ESGC score, which discounts for ESG controversies impacting the firm. The final rating is issued both in points from 0-100 and in letter grades from D- to A+, with a higher score or grade indicating better ESG performance. ESG data and scores are recalculated on an ongoing basis to align with corporate reporting patterns (Refinitiv, 2022b, p. 3 f.).

The next chapter delves into the issue of disagreement among ESG rating agencies. In particular, the extent to which ESG rating agencies disagree and the reasons for their disagreement. The chapter aims to build a theoretical foundation for the independent variable of this master thesis.

4.3. Disagreement among ESG Rating Agencies

ESG rating agencies can disagree significantly with respect to their ESG ratings. In a recent study, Berg et al. (2022) examine the disagreement between the ESG ratings of five major ESG rating agencies (KLD, Sustainalytics, Moody, Refinitiv and S&P Global). They find an average correlation of only 54% between the ESG ratings (see Table 1), which is surprising since these ESG ratings are supposed to measure the same risk construct. At the pillar level, the disagreement is even higher with correlations of 0.53, 0.42, and 0.30 for E, S, and G, respectively. ESG rating agencies appear to disagree the most on governance issues, with some ESG rating agencies even exhibiting negative correlations. The negative correlations indicate extreme disagreement among ESG rating agencies. Firms that were considered to have good ESG performance by one ESG rating agency, were considered to have bad ESG performance by the other ESG rating agency. The results indicate that the information investors receive from ESG rating agencies is relatively noisy.

Other studies support the notion that there is a significant disagreement among ESG rating agencies. Prall (2021) analyses the correlations between six major ESG rating agencies (MSCI, S&P, Sustainalytics, CDP, ISS and Bloomberg). He finds even lower correlations between those ESG rating agencies, with an average correlation of just 35%. MSCI's correlation with both Sustainalytics and S&P is below 50% (see Table 17 in the annex). The rest of the correlations range from 0.74 (between S&P and Bloomberg) to 0.07 (between ISS and CDP). State Street Global Advisors (2019, p. 2) assesses cross-sectional correlations between four major rating agencies (Sustainalytics, MSCI, RobeccoSAM and Bloomberg). The results show an average correlation of 60%. The correlation between Sustainalytics and MSCI is only 53% (see Table 18 in the annex), which is consistent with the findings of Prall (2021). Boffo and Patalano (2020, p. 28) examine the ESG rating variation among three major ESG rating agencies (Bloomberg, MSCI, and Refinitiv) for the components of the S&P 500 and STOXX 600 indices. They find large differences, with an average R-squared of 0.21 for the S&P 500 and 0.18 for the STOXX 600. From a correlation perspective, these values correspond to 46% and 42% for the two indices, respectively.

In another analysis, Boffo and Patalano (2020, p. 29) compare the disagreement between ESG rating agencies and credit rating issuers. For this purpose, they selected listed firms by largest market capitalization to represent various industries. The results show that ESG ratings agencies disagree significantly in their ESG ratings, while credit rating issuers mostly agree (see Figure 1). Berg et al. (2022, p. 6 f.) even report a correlation between credit ratings of 99%. Prall (2021) find that the credit ratings for the firms in their

Table 1: Correlations between ESG ratings (Source: Berg et al. (2022, p. 30))

	KL SA	KL MO	KL SP	KL RE	KL MS	SA MO	SA SP	SA RE	SA MS	MO SP	MO RE	MO MS	SP RE	SP MS	RE MS	Average
ESG	0.53	0.49	0.44	0.42	0.53	0.71	0.67	0.67	0.46	0.7	0.69	0.42	0.62	0.38	0.38	0.54
E	0.59	0.55	0.54	0.54	0.37	0.68	0.66	0.64	0.37	0.73	0.66	0.35	0.7	0.29	0.23	0.53
S	0.31	0.33	0.21	0.22	0.41	0.58	0.55	0.55	0.27	0.68	0.66	0.28	0.65	0.26	0.27	0.42
G	0.02	0.01	-0.01	-0.05	0.16	0.54	0.51	0.49	0.16	0.76	0.76	0.14	0.79	0.11	0.07	0.30

Note: Correlations between ESG ratings at the aggregate rating level (ESG) and at the level of the environmental dimension (E), the social dimension (S), and the governance dimension (G). SA, SP, MO, RE, KL, and MS are short for Sustainalytics, S&P Global, Moody's ESG, Refinitiv, KLD, and MSCI, respectively.

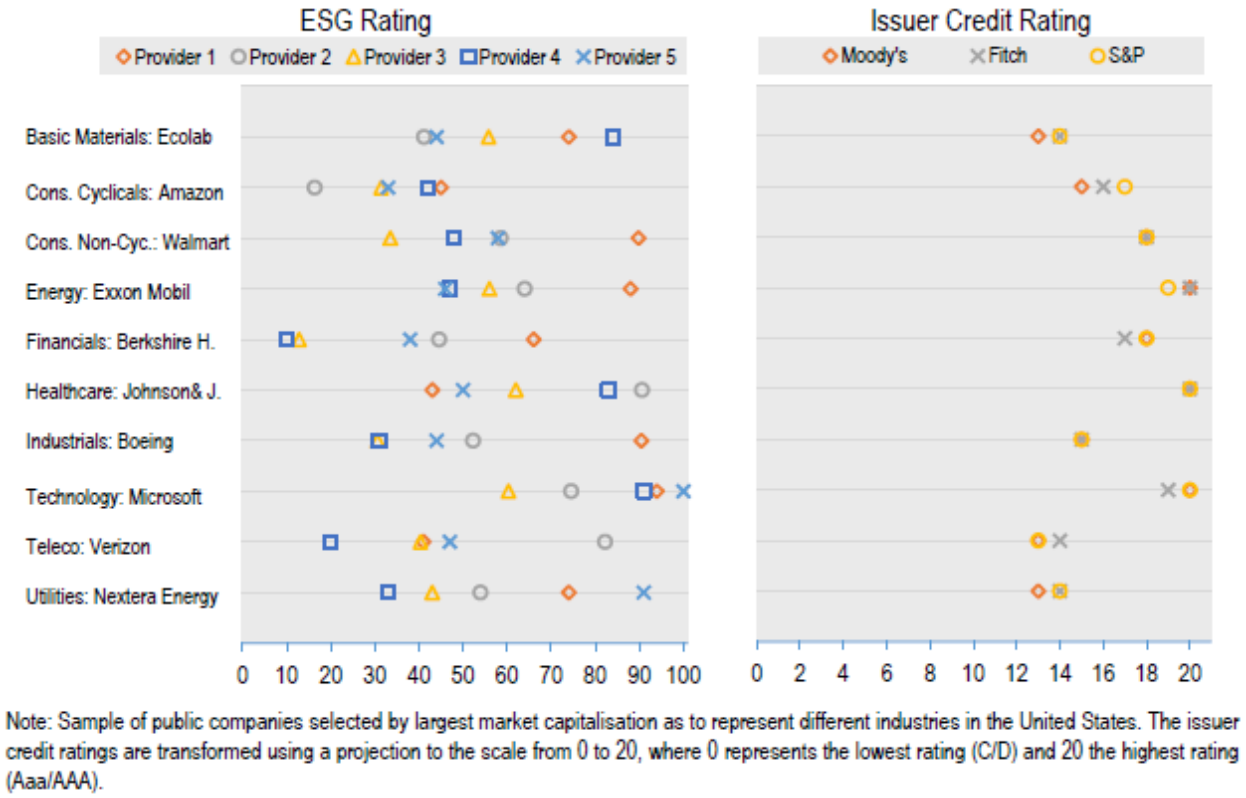
sample have a correlation between 94% and 96%. Therefore, the disagreement seems to be unique to non-financial rating agencies.

This raises the question why ESG rating agencies disagree that much. As mentioned in the previous chapter, ESG rating agencies use very different rating methodologies to collect, measure and analyze ESG information. The disagreement can be mostly attributed to different ESG rating methodologies. Because ESG rating agencies compete with each other for market share, there is no single approach to ESG rating methodologies. Each ESG rating agency uses its own proprietary methodology to differentiate itself from their peers and to meet investors' needs (Brackley et al., 2022). In addition, ESG rating methodologies are often not fully transparent (Doyle, 2018, p. 8). Berg et al. (2022) seek to understand which factors contribute to the disagreement among ESG rating agencies. They deconstruct ESG ratings into three factors: scope (the attributes that the ESG rating agencies attempt to measure), measurement (the measures used to assess the attributes), and weighting (the relative importance assigned to the attributes). They find that the majority of disagreement between ESG rating agencies can be attributed to differences in measurement (56%) and scope (38%), with weighting differences accounting for only 6% of the disagreement. The one exception to the study is MSCI, where the scope, rather than the measurement, accounts for most of the disagreement due to the firm-specific weights (Berg et al., 2022, p. 16 f.).

At the scope level, ESG rating agencies differ in the amount and type of input variables. While several ESG rating agencies use ESG frameworks, such as GRI, SASB, and TCFD to select input variables, others do not. Input variables are also selected to some degree based on data availability to ensure that each indicator can be accurately measured over time. In cases where firms do not provide direct information, approximations are used, which may or may not be accurate. Perhaps contrary to expectations, Christensen et al. (2021, p. 5 f.) find that increased firm disclosure does not lead to more consistent ESG ratings. Instead, they find that it actually increases the disagreement between ESG rating agencies. This is because the subjective nature of ESG information allows for different interpretations of the disclosed information, leading to greater disagreement among ESG rating agencies. Input variables can also differ significantly across industries or firms to account for financial materiality (Boffo & Patalano, 2020, p 31). ESG rating agencies may also replace input variables through time, making it difficult to compare ESG ratings over time or even leading to changes in past ESG ratings (Escrig-Olmedo et al., 2019, p. 14).

At the measurement level, disagreement between ESG rating agencies can arise due to differences in the interpretation of ESG information. For instance, ESG rating agencies use expert judgement to determine which input factors are material for various industries, how to interpret various input factors, or how to handle data gaps (Boffo and Patalano 2020, p. 31; Kotsantonis and Serafeim 2019, p. 54). Berg et al. (2022, p. 18) find that the ESG rating agencies' assessment of a firm in individual categories can influence their overall view of the firm, a phenomenon they called the rater effect. When a rater had a positive view of a firm's particular indicator, they were more likely to have a positive view of the firm's other indicators as well. Berg et al. (2022, p. 17) further find that certain categories are more prone to disagreement. ESG rating agencies mostly disagree on climate risk management, product safety, corporate governance, corruption and environmental management systems. Other categories, such as environmental fines, clinical trials, employee turnover, HIV programs and non-greenhouse gas air emissions are less prone to disagreement. Another factor that influences the interpretation of ESG information is experience. Many investors criticize the insufficient seniority and tenure of research analysts who develop ESG ratings, stating that research teams are stretched too thin and do not have a deep enough understanding of the issues and sectors (The SustainAbility Institute, 2020, p. 29). In any case, the level of experience affects the quality of ESG ratings and thus the disagreement between ESG rating agencies.

Finally, there are weights. ESG rating agencies often assign different weights to different input variables and categories. These weights can be either determined by expert judgment or based on quantitative data-driven approaches (Boffo & Patalano, 2020, p 31). Input variables and categories that have a greater impact on the firm's financial performance often receive a higher weighting (Larcker et al., 2022, p. 4). Since weights are assigned by the individual ESG rating agencies, there may be differences in weightings



Note: Sample of public companies selected by largest market capitalisation as to represent different industries in the United States. The issuer credit ratings are transformed using a projection to the scale from 0 to 20, where 0 represents the lowest rating (C/D) and 20 the highest rating (Aaa/AAA).

Figure 1: Comparison of disagreement between ESG ratings and credit ratings (Source: Boffo and Patalano (2020, p. 29))

that can lead to disagreement between ESG rating agencies.

The disagreement between ESG rating agencies can be both unintentionally and intentionally. Unintentional disagreement in ESG ratings often occurs at the level of specific input factors or data points. Unintentional divergence in ESG ratings may occur when different raters evaluating the same firm have different access to data or interpret the same information differently, resulting in divergent conclusions about the firm's ESG performance. Intentional disagreement in ESG ratings typically occurs at the composite ESG score level, and is the result of the rater's comprehensive analysis of the firm's ESG performance based on its own methodology. This disagreement reflects the differing perspectives and approaches used by the different ESG rating agencies in evaluating a firm's ESG performance (Brackley et al., 2022).

The disagreement between ESG rating agencies, which is often caused by the lack of consistency and standardization in rating methodologies, can limit the usefulness of ESG ratings in providing reliable and meaningful information about a firm's long-term resilience and non-financial performance (Brackley et al., 2022). Without a consistent and standardized approach to ESG ratings, it can be difficult to compare and evaluate the ESG performance of different firms, making it challenging to use ESG ratings as a tool for informed decision-making (Larcker et al., 2022, p. 6). However, while greater consistency in ESG ratings may be desirable in terms of providing more reliable and meaningful information about a firm's performance, it is not clear whether investors necessarily want greater consistency in rating methodologies. Greater regulation of ESG ratings may help to standardize the information input and rating process, resulting in more consistent ratings and reducing the disagreement between ESG rating agencies. On the one hand, greater consistency may reduce the amount of conflicting or contradictory ESG ratings, making it easier for investors to compare and evaluate the ESG performance of different firms. On the other hand, the inclusion of multiple perspectives and approaches in the ESG rating process may provide a more comprehensive and nuanced view of a firm's performance, and may be seen as a positive characteristic by some investors (Brackley et al. 2022; The SustainAbility Institute 2020, p. 44 f.).

The next chapter focuses on the development of the hypothesis. This chapter aims to provide a theoretical framework that can be used to make predictions about the association between ESG disagreement and the dispersion of analysts' forecasts.

5. Hypothesis Development: Influence of ESG Rating Disagreement on Analyst Forecast Dispersion

In this master thesis, I seek to understand the relationship between ESG rating disagreement and analyst forecast dispersion. ESG ratings are measures of a firm's performance in relation to ESG criteria. These ratings aim to measure

a firm's exposure to ESG risks and opportunities, and how those risks and opportunities may impact the firm's financial performance (MSCI, 2022b, p. 3). ESG rating disagreement refers to the degree of variation in these ratings among different ESG rating agencies. Analyst forecast dispersion refers to the degree of disagreement among analysts in their forecasts of a firm's future EPS performance. In the accounting and finance literature, analyst forecast dispersion is widely recognized as an important measure, and is often used as a proxy for the uncertainty and the divergence in analysts' beliefs and the lack of consensus or agreement (Barry and Jennings 1992, p. 172; Abarbanell et al. 1995, p. 32; Barron et al. 2010, p. 422).

There is ongoing debate in the literature about why ESG rating agencies may disagree in their ratings of a firm's performance and how ESG information is relevant to market participants. Christensen et al. (2021, p. 4-6) examine whether a firm's ESG disclosure impacts the disagreement between ESG rating agencies. They find that greater ESG disclosure leads to greater ESG rating disagreement. They further find that ESG disagreement is associated with higher stock return volatility and larger absolute price movements, and is therefore relevant to market participants. Krueger et al. (2021, p. 35) study how mandatory ESG disclosure affects the dispersion of analysts' earnings forecasts. They find that as mandatory ESG disclosure improves, analyst earning forecasts become less dispersed. They also find that mandatory ESG disclosure significantly reduces the amount of negative ESG incidents in a firm-year (Krueger et al., 2021, p. 49). Cho et al. (2013, p. 81 f.) investigate whether CSR performance reduces the bid-ask spread, a proxy for information asymmetry. They find that both positive and negative CSR performance seem to reduce information asymmetry. Information asymmetry itself is often interpreted as a constituent of uncertainty (Barron et al., 2010, p. 333). Having said that, the literature on the relationship between ratings and analyst forecast dispersion is scarce. Avramov et al. (2009, p. 85) examine a dispersion-based trading strategy. They find that a portfolio strategy based on buying low dispersion stocks and selling high dispersion stocks yields a statistically significant return. They further find that recent credit rating downgrades lead to higher analyst forecast dispersion (Avramov et al., 2009, p. 99 f.). There are even fewer studies when it comes to the relationship between ESG rating disagreement and analyst forecast dispersion. In fact, during my research I were only able to find one study that addressed this relationship. Kimbrough et al. (2022, p. 48) examine whether ESG rating disagreement is associated with disagreement among market participants. They find that ESG rating disagreement is positively associated with analyst forecast dispersion, bid-ask spread and future stock return volatility. Though, the relationship between ESG rating disagreement and analyst forecast dispersion is only statistically significant at the 10% level, indicating a weak link between the two variables and that the relationship may not be causal. Also, Kimbrough et al. (2022) analyzed the relationship between ESG rating disagreement and analyst forecast dispersion in the US. Therefore, further research may be needed to confirm or refute the relationship between the two variables. This master thesis aims to fill this research gap by conducting an empirical analysis on the relationship between ESG rating disagreement and analyst forecast dispersion in an international setting (Kimbrough et al., 2022, p. 48). Since research on the association between ESG disagreement and analyst forecast dispersion is sparse, the hypothesis development is discussed in more detail. Different arguments for a positive, negative and no association are presented. A decision is then made in favor of one direction or the other based on the strongest arguments.

There are several arguments that could be made in favor of a positive relationship between ESG rating dispersion and analyst forecast dispersion. First, if analysts use different ESG ratings, this could lead to differences in their EPS forecasts, as each ESG rating agency provides different information and perspectives. The access to ESG ratings can be costly, with institutional investors on average spending on $ 487,000 per year on ESG ratings, data and consultants (The SustainAbility Institute, 2022, p. 5). This means that some analysts may not have the resources to access paid ESG ratings services or may choose to use fewer of them in their evaluations. This could lead to differences in the ESG ratings used by analysts, resulting in variations in their forecasts. Additionally, the selection of ESG ratings by individual analysts may be a factor, as some rating agencies are more likely to disagree with others (see Table 17 and 18 in the annex). According to Capital Group (2021, p. 29), the majority of investors use between two and five different ESG ratings (57%), while some use only one (24%) or none at all (7%). This means that it is possible that analysts are not using the same ESG rating agencies in their assessments, which could contribute to the dispersion in their forecasts. In the future, it is expected that the number of ESG ratings used by investors will increase, which may lead to a decrease in the effect of ESG disagreement on analyst forecast dispersion as the variations in ESG ratings are averaged out.

Second, even though analysts may use the same ESG ratings, their interpretations and resulting EPS forecasts can vary significantly. This is because some analysts may simply view ESG ratings as a form of box-checking exercise and do not delve deeper into how ESG rating agencies arrive at their ESG ratings (The SustainAbility Institute, 2020, p. 31). Others may use ESG ratings as a starting point for further research, scrutinizing the measurement, scope, and weights of the ratings in their analysis. High levels of disagreement among ESG rating agencies in particular can be seen as a reason for a more in-depth analysis (Boffo & Patalano, 2020, p. 29 f.). As a result, analysts may develop different private knowledge about ESG ratings, leading to dispersion in analyst EPS forecasts. This view is consistent with Lang and Lundholm (1996, p. 471 f.), who argues that that as public information becomes less informative, analysts place more emphasis on their private information. It is also consistent with Behn et al. (2008, p. 330) who argues that greater dispersion may reflect a lack of agreement among analysts, potentially due to some analysts' inability or reluctance to

fully and objectively gather and interpret ESG-related information.

Third, analysts may disagree about whether ESG ratings actually reflect a firm's non-financial performance, as there seems to be no consensus even among ESG rating agencies. Assuming that ESG rating agencies observe the same firm-disclosed ESG information, and rate firms based on their non-financial risks and opportunities, there should be no dispersion in ESG ratings. But, ESG rating agencies seem to be not sure what constitutes as good or bad ESG performance, resulting in widely divergent ESG ratings (Boffo & Patalano, 2020, p. 64 f.). This raises questions about the credibility and reliability of these ratings as a measure of firms' non-financial performance (Larcker et al., 2022, p. 6). As a result, analysts rely more on private information in addition to ESG ratings (Lang & Lundholm, 1996, p. 471 f.), leading to divergent EPS forecasts.

There are also two arguments that could be made in favor of a negative relationship between ESG rating dispersion and analyst forecast dispersion. First, ESG rating dispersion could serve as a proxy for the disclosure of heterogenous ESG information, which in turn can lead to a reduction in analyst forecast dispersion. ESG rating agencies act as information intermediaries by gathering, aggregating and evaluating a firm's public non-financial information. Some ESG rating agencies even conduct their own surveys, therefore producing and facilitating their own disclosure of ESG information (Scalet & Kelly, 2010, p. 71). Under the premise that analyst forecast dispersion reflects the amount of information commonly available to analysts, forecast dispersion should decrease with more ESG information being available (Han & Manry, 2000, p. 119). This is because if analysts share a common forecasting model and observe the same ESG information but have different private information, they will attach less weight to their private information as the informativeness of ESG information increases, thereby reducing forecast dispersion (Lang & Lundholm, 1996, p. 471). This view is consistent with Krueger et al. (2021, p. 9) who argues that as more and better ESG information is made available, the diversity of opinions may decrease, and EPS forecast dispersion should decrease. Next to the quantity of disclosure, the quality also seems to be important. Swaminathan (1991, p. 40) find that forecast dispersion decreases following the release of newly mandated segment information by the SEC. Dechow et al. (1996, p. 3) find that forecast dispersion increases following alleged violation of generally accepted accounting principles. Because ESG information is largely unstandardized, frequently unstructured, difficult to compare and tends to be more subjective than financial disclosures (Sipiczki, 2022, p. 6), one could argue that through the aggregation and evaluation of unstandardized and unstructured ESG information, ESG rating agencies increase the quality of ESG disclosures, thereby reducing analyst forecast dispersion.

Second, ESG rating disagreement may reflects different perspectives and approaches of ESG rating agencies, allowing for a more comprehensive and nuanced understanding of a firm's ESG performance, and thus reducing the dispersion of analysts' forecasts. When ESG rating agencies have different perspectives and approaches to evaluating a firm's ESG performance, it leads to a more comprehensive and nuanced understanding of the firm. This is because the ESG ratings become more dispersed, meaning they reflect a wider range of viewpoints and a greater amount of underlying data (Scalet and Kelly 2010, p. 72; The SustainAbility Institute 2020, p. 44 f.). As a result, analysts have access to more information and can form a more informed opinion about a firm's financial prospects. This ultimately leads to a decrease in forecast dispersion and increased agreement among analysts. For this to hold true, though, analysts would have to have access to the same ESG ratings and interpret them in the same way (Lang & Lundholm, 1996, p. 471 f.).

In addition, there are several arguments why there may not be a significant relationship between ESG rating dispersion and analyst forecast dispersion. First, ESG disagreement may not have an effect on analyst forecast dispersion if ESG ratings reflect a firm's long-term ESG performance, while analyst forecasts reflect a firm's short-term profitability. In this view, ESG ratings provide analysts with information about a firm's long term risks and opportunities (Boffo & Patalano, 2020, p. 14). For example, a poor environmental performance can lead to negative consequences such as fines, legal action, and damage to a firm's reputation, which in turn can affect financial performance. Whereas, a good environmental performance can improve a firm's reputation and mitigates the risk of regulatory scrutiny (Henisz et al., 2019, p. 3-8). However, it is difficult to predict when these ESG risks will materialize in the future. In contrast, analyst EPS forecasts are projections of a firm's short-term financial performance, with a time horizon typically limited to the next quarter or fiscal year. Therefore, most ESG risks are unlikely to be relevant to analysts' EPS forecasts and may not be used when making EPS forecasts. Still, some ESG rating agencies include controversies into their ESG ratings. Controversies are short-term reputational risks that arise from negative media attention (Boffo & Patalano, 2020, p. 30). Because these controversies affect a firm's short-term performance, analyst may consider ESG ratings when making their EPS forecasts. As a result, the relationship between ESG rating disagreement and analyst forecast dispersion depends on whether ESG ratings reflect both short-term and long-term ESG performance.

Second, ESG ratings dispersion may not affect analyst forecast dispersion due to a lack in the transparency of ESG ratings. ESG rating agencies tend to not fully disclose their ESG rating methodologies. Investors lack an clear understanding about which metrics, inputs and weights ESG rating agencies use in their evaluation, as well as the degree of subjectivity that in their assessments (Brackley et al., 2022). This lack of transparency makes it difficult for analysts to use ESG ratings as a reliable source of information to inform their earnings forecasts. As a result, analysts may use other sources of ESG information beyond ESG ratings to inform their earnings forecasts such as firm-provided disclosures, market and industry trends and specific news and events.

This reliance on other sources of ESG information reduces the significance of ESG ratings and their disagreement, causing analysts to ignore ESG ratings.

Third, the dispersion of ESG ratings may not be relevant to analysts' forecasts because of the backward-looking data used in ESG ratings. ESG rating agencies mostly use publicly available information to access a firm's ESG performance. Therefore, they can produce an accurate assessment of a firm's past ESG performance. But, analysts are interested in forecasting a firm's future financial performance (The SustainAbility Institute, 2020, p. 28). Sheng and Thevenot (2012, p. 21) argue that analysts' EPS forecasts represent market participants' expectations of a firm's future earnings prior to the release of accounting data. Past ESG information may already be prized in by the market (Malkiel & Fama, 1970, p. 383). Also, past performance is not necessarily a reliable indicator of future performance, which is why analysts use estimates and correct their forecasts on an ongoing basis (Capstaff et al., 1995, p. 74). Thus, ESG ratings may be of limited use for future investment decisions and are therefore not considered by analysts in their forecasts. As a result, there would be no significant relationship between the dispersion of ESG ratings and the dispersion of analysts' forecasts.

Having considered all the arguments, I believe that analyst forecast dispersion is driven by differences in the interpretation and use of ESG ratings. Accordingly, a positive association between ESG rating dispersion and analyst forecast dispersion is considered the most likely hypothesis. Therefore, I hypothesize:

> **H1**: ESG rating disagreement is positively associated with analyst forecast dispersion

This means that as the dispersion between ESG ratings increases, the dispersion in analyst forecast also increases. To test whether there is a positive relationship between the dispersion of ESG ratings and the dispersion of analyst forecasts, I conduct an empirical analysis.

6. Empirical Study

6.1. Sample

I start with an initial sample of 7,186 global public firms obtained from Refinitiv Eikon. The firms are constituents of the Market WD index. The initial sample consists of 71,860 firm-year observations ranging from 2012 to 2022. The necessary firm data and the in-house ESG ratings were collected from Refinitiv Eikon. The time period of 10 years is chosen so that the earnings volatility of the last 5 years can be calculated correspondingly for each ESG rating observation. In a first step, I make sure that the sample does not contain duplicates, i.e., does not contain more than one observation belonging to the same firm-year. In a second step, I ensure that all firm-year observations are distinctly attributable to a single firm and a single fiscal year. Then, with the exception of the disagreement between ESG rating agencies, I calculate all variables required for the empirical analysis and remove missing observations from the dataset.

This subsample is then used to collect the respective other ESG ratings. In total, I hand-collect ESG ratings from 4 prominent ESG rating providers: MSCI, S&P, ISS, Sustainalytics. When a ESG rating agency released multiple ESG ratings for a given firm year, I collected the last ESG rating provided for a given year. The ESG ratings collected vary in data availability. For some ESG ratings, such as Sustainalytics and ISS, only the latest ESG ratings for 2022 are available, while MSCI and S&P, for example, provide ESG ratings covering a period from 2018 to 2022. Also, not all ESG rating agencies publish their corresponding E/S/G pillar scores to the ratings. In order to obtain a sufficiently large data basis, S&P and ISS were included in the empirical analysis. The initial intention was to include only MSCI, Sustainalytics and Refinitiv Eikon. However, the data basis would then have been too small. For the empirical analysis, a total of 9,577 ESG ratings from both MSCI and S&P were accessed, resulting in 3,785 and 3,888 ESG ratings respectively. In the case of Sustainalytics and ISS, 385 ESG ratings were accessed, resulting in 329 and 284 ratings, respectively. In addition, the pillar scores for S&P and Sustainalytics for the year 2022 were also collected.

After collecting the ESG ratings, the two datasets are merged and subsequently adjusted for missing values in the ESG disagreement calculation. The final sample consists of 3,968 firm-year observations ranging from 2018 to 2022. Table 19 in the annex shows the respective sample selection procedure. As mentioned previously, the sample is an international sample. All available country observations were collected, with the exception of the US. In fact, the final sample consists of 54 unique countries. The three largest positions are Japan, India and the United Kingdom, which account for 19.5%, 8.1% and 5.2% of the sample, respectively. Table 20 in the annex shows the composition of the sample by countries. The sample differs from the study by Kimbrough et al. (2022) in two important ways. First, Kimbrough et al. (2022, p. 5) focus only on firms in the US due to the voluntary nature of ESG information reporting, whereas our sample includes all countries excluding the US[1]. Second, Kimbrough et al. (2022, p. 2) collect ESG rating information from KLD (now MSCI), ASSET4 (now Refinitiv Eikon), and Vigeo Eiris (now Moody's). Therefore, this study differs from Kimbrough in that the type and quantity of different ESG ratings and the country choice differs. The next chapter addresses the research design of this empirical study.

6.2. Research Design

To test the hypothesis whether there is a positive association between analyst forecast dispersion and ESG disagreement, I perform an empirical analysis based on an OLS regression,

$$AF\_DISP_{i,t} = \beta_0 + \beta_1 ESG\_Disagreement_{i,t} + \beta_k Controls_{i,t} + \varepsilon_{i,t} \quad (1)$$

[1] All countries refers to all the countries included in the Market WD index.

where $AF\_DISP_{i,t}$ is the dependent variable, $ESG\_Disagreement_{i,t}$ the independent variable, $Controls_{i,t}$ the control variables and $\varepsilon_{i,t}$ the error term. Table 21 in the annex reports all variables used in the regression analysis. $AF\_DISP_{i,t}$ refers to the relative dispersion between analysts' forecasts. It is calculated as the natural logarithm of the standard deviation of analysts' forecast dispersion of annual EPS scaled by the absolute value of the mean analysts' forecast for firm i in year t. The absolute value is important to be mathematically correct since in a natural logarithm one cannot divide by a negative number. Otherwise, observations would be lost during the analysis. Kimbrough et al. (2022, p. 39) and Cui et al. (2018, p. 21) scale analyst dispersion using the absolute value of the mean. However, while Kimbrough et al. (2022) use the natural logarithm, Cui et al. (2018) do not. In addition, Krueger et al. (2021, p. 51) define analyst dispersion as the standard deviation of analysts' forecasts divided by the stock price for firm i in year t. Initially, I wanted to use the standard deviation of analysts' forecasts to calculate analyst dispersion ($AF\_Dispersion\_0_{i,t}$). But, as can be seen in the histogram in Figure 5 in the annex, the observations are not-normally distributed using this measure. For this reason, I used the natural logarithm to transform analyst forecast dispersion. After that, the sample observations for analyst dispersion are normally distributed as indicated by the bell curve (see Figure 6). $ESG\_Disagreement_{i,t}$ is the variable of interest. It is calculated as the natural logarithm of the standard deviation of ESG ratings scaled by the absolute value of the mean ESG forecast for firm i in year t. This measure is used to make $AF\_Dispersion_{i,t}$ and $ESG\_Disagreement_{i,t}$ comparable. In contrast, Kimbrough et al. (2022, p. 39) use the absolute value of the difference between the percentile rank of ESG ratings as a measure of ESG dispersion. They also use the standard deviation of the percentile ranks of ESG ratings as a measure of ESG dispersion in their study, but not when examining the influence on analyst forecast dispersion (Kimbrough et al., 2022, p. 48). Christensen et al. (2021, p. 39) use ESG disagreement as the dependent variable and calculate it using the standard deviation of ESG ratings. In the empirical analysis, the standard deviation of ESG ratings is calculated in such a way that if an ESG rating is missing, the standard deviation is still calculated for the available ESG ratings. Apart from this, at least three ESG ratings are required. To arrive at $ESG\_Disagreement_{i,t}$, ESG ratings themselves must first be made comparable. Each ESG rating provider uses its own rating scale, which makes it difficult to compare ESG ratings. Refinitiv (2022b, p. 3) and S&P Global (2022, p. 3) use a percentile rank scores between 0 and 100, where 100 represents the best score. Sustainalytics (2020, p. 7) also uses a percentile rank score. But, the percentile ranks range from 1 to 100, with 0 being the best and 100 the worst. MSCI (2022b, p. 12) uses a letter-based rating system with 12 categories, where AAA represents the best score and CCC the worst. ISS also uses a letter-based rating system. However, ISS ESG (2022, p. 2) uses only 7 letters, with D- representing the worst and A+ representing the best score .To make the ESG ratings of Refinitiv Eikon, MSCI, S&P, ISS and Sustainalytics comparable, I first change the direction of Sustainalytics' ESG score so that 100 represents the best and 1 the worst. Then I standardize Sustainalytics' ESG score so that 0 represents the worst score. Then I divide the ESG scores of the three ESG providers by 10 to arrive at a 10-point rating rank scale, which seems more appropriate given the lower number of score grades from MSCI and ISS. After that, I convert the letter-based scores from MSCI and ISS into numeric scores. Since one letter equals zero, I divide the highest possible score ten by $n-1$ to arrive at the respective numerical scores for MSCI and ISS (See Equation 2).

$$0 + \frac{10}{(n-1)} = numeric\ score\ rank \tag{2}$$

I also construct three alternative measures of ESG disagreement. The first alternative measure is $ESG\_Disagreement\_3$. Similar to $ESG\_Disagreement$, it is computed as the natural logarithm of the standard deviation of ESG ratings scaled by the absolute value of the mean ESG forecast for firm i in year t. The individual ESG ratings are also made comparable in the same way as for $ESG\_disagreement$. The difference is that for $ESG\_Disagreement\_3$ only the ESG ratings of Refinitiv Eikon, MSCI and S&P are used to calculate the standard deviation. Similarly, $ESG\_Disagreement\_4$ is calculated using the four ESG ratings from Refinitiv Eikon, MSCI, S&P and Sustainalytics. $ESG\_Disagreement\_5$ uses all five ESG ratings. If an ESG rating is not available in a particular firm year, the alternative measures are not calculated for this particular firm year. It is therefore required that all ESG ratings necessary for the calculation are available.

In addition, I construct three measures to analyse the disagreement among ESG rating agencies on the E/S/G pillar scores. The measures only include the pillar scores of Refinitiv Eikon, S&P and Sustainalytics, as the other pillar scores are not publicly available free of charge. Similar to ESG_Disagreement, E/S/G_disagreement it is computed as the natural logarithm of the standard deviation of the E/S/G pillar scores scaled by the absolute value of the mean E/S/G forecast for firm i in year t. E_Disagreement captures the disagreement among ESG rating agencies about environmental issues. S_Disagreement captures the disagreement While S_Disagreement captures the disagreement between ESG rating agencies on social issues and G_Disagreement on governance issues. One issue is to make the pillar scores comparable. Refinitiv and S&P create pillar scores and subsequently weight them to arrive at their ESG ratings. The pillar scores of Refinitiv and S&P are directly comparable. This is because the subsequent weighting does not affect the individual pillar scores. In the case of Sustainalytics, the sum of the individual pillar scores equals the final ESG rating. It is not entirely clear from Sustainalytics' rating methodology how the individual E/S/G pillars are weighted. To perform an empirical analysis, an equal weighting is assumed. The pillar scores from Sustainalytics are therefore multiplied by three to arrive at a comparable pillar score.

To control for the influence of other variables, I include several control variables in my empirical analysis. The control variables are chosen based on similar previous studies. Controls consists of firm size, book-to-market-ratio, analyst following, earnings surprise, forecast horizon, earnings volatility, indicator variable for negative earnings, leverage and Zmijewski financial distress score (Christensen et al. 2021, p. 40; Behn et al. 2008, p. 333 f. Hope 2003, p. 25; Kimbrough et al. 2022, p. 38). The calculations for the control variables are given in Table 21 in the annex. Firm size is included because large firms would be expected to have a smaller dispersion (Behn et al., 2008, p. 333). Analyst following is included based on Lang and Lundholm (1996, p. 482), who find a positive association between analyst following and forecast characteristics. Earnings surprise is also based on Lang and Lundholm (1996, p. 489), who find that larger changes in earnings are related to less accurate forecasts. Forecast horizon is considered based on Chopra (1998, p. 37), who finds that a forecast further away from to the actual earnings announcement date is less accurate and more dispersed than a forecast closer to the announcement date. However, because many firm-year observations are missing to calculate the variable, the control variable is ultimately not included in the study. Earnings volatility is included based on Kross et al. (1990, p. 465) who find that firms with large historical earnings variations have less accurate analyst's earnings forecasts. Variability in earnings should increase the difficulty of forecasting, resulting in larger dispersion. The indicator variable for negative earnings, leverage and Zmijewski financial distress score are included to control for uncertainties arising from strained financial conditions and bankruptcy risk. The indicator variable for negative earnings is included based on Hwang et al. (2014, p. 29) who find that analysts' forecasts for firms with negative earnings are on average less accurate than for firms with positive earnings. Leverage is included based on Hope (2003, p. 11) who mentions that highly levered firms tend to have more variable earnings. Zmijewski (1984, p. 65-69)'s financial distress score is included based on Behn et al. (2008, p. 333) who note that financially distressed firms tend to have less accurate forecasts. The book-to-market ratio is included based on the Kimbrough et al. (2022, p. 19) to control for growth opportunities related to ESG. In addition, I further include industry and year fixed effects. The variables are winzorized at both tails at the 1% level.

7. Empirical Results

7.1. Descriptive Statistics

Table 2 reports the descriptive statistics for the individual ESG ratings. As can be seen in Table 2, Refinitiv, MSCI and S&P are represented in the sample with around 4000 ESG ratings each, while ISS and Sustainalytics are only represented with just around 300 ratings.

Furthermore, it can be seen that the largest observation of ISS has a value of 6.36, and not close to ten. This is due to the fact that no ESG ratings better than B have been assigned to the firms in the sample. Accordingly, no ESG ratings from ISS for firms with excellent ESG performance are represented in the sample. In addition, it can be seen that the smallest observation for Sustainalytics has a value of 4.63, and not close to zero. Thus, Sustainalytics is distorted for firms with particularly poor ESG performance. This is because Sustainalytics assigns firms to the worst category at a value above 40. The assigned nominal value, though, goes beyond 40. To avoid distortions and make Sustainalytics comparable, one could set the maximum observed value as the upper limit and then adjust the other ESG ratings accordingly. But, due to the subordinate role of Sutainalytics in the sample and other robustness checks, this approach was not applied here. Still, it is important to be aware of this bias for the further course of this empirical analysis. It is also noticeable that the ESG ratings of S&P and ISS have a relatively low mean of 3.84 and 2.91. Together with the also low median values, this indicates that S&P and ISS generally assign lower ESG ratings than Refinitiv, MSCI and Sustainalytics. It is also worth noting that the standard deviation of MSCI and S&P with 2.68 and 2.48 are higher than those of the other ESG rating providers. This indicates that the ESG ratings of MSCI and S&P are more dispersed around the mean. Thus, a greater variability in ESG ratings.

Table 3 shows the correlation between the ESG ratings of different ESG rating providers. The correlations between the ESG ratings are low. This is consistent with the observations of Prall and State Street Global Advisors (see Table 18 and 19). Hence, ESG rating providers generally do not agree about the ESG performance of firms. Therefore, resulting in high levels of disagreement among ESG rating agencies. The highest levels of disagreement are found between Sustainalytics and other ESG rating providers. Yet, some of the correlations are not empirically significant at the 1% level. The highest levels of agreement are found between Refinitiv and S&P, and S&P and ISS with 0.55 and 0.55, respectively.

Table 4 and 5 present the descriptive statistics for the empirical analysis. Table 4 shows the descriptive statistics for analyst forecast dispersion and ESG disagreement before the transformation with the natural logarithm. Both variables are calculated as the standard deviation (See Table 21). The mean and median of AF_DISP_0 are 49.38 (0.43). The standard deviation of AF_DISP_0 is 342.99. These statistics indicate that there are substantial variations in forecasts made by financial analysts.

The reason why I transform analyst forecast dispersion is that the variable is highly dispersed around the mean, highly skewed, and exhibits a high positive kurtosis. All of this can be problematic for accuracy of the hypothesis test. First, the standard deviation is greater than the mean. Hence the coefficient of variation[2] is more than one. This means that analyst forecast dispersion exhibits a great degree of relative variability. A great degree of variability in the data set is

[2] The coefficient of variation is defined as the ratio of the standard deviation to the mean and is a standardized measure of dispersion.

Table 2: Descriptive Statistics of ESG ratings (Source: Own illustration)

Variable	N	Mean	SD	25%	Median	75%	Min	Max
Refinitiv	3,968	5.87	1.76	4.75	6.02	7.25	0.072	9.424
MSCI	3,785	5.65	2.68	3.33	6.67	8.33	0	10
S&P	3,888	3.84	2.48	1.8	3.2	6.0	0	9.3
ISS	284	2.91	1.26	1.82	2.73	3.64	0.91	6.36
Sustainalytics	329	7.59	0.84	7.07	7.68	8.26	4.63	9.39

Table 3: Correlations of ESG ratings (Source: Own illustration)

	(1)	(2)	(3)	(4)	(5)
(1) Refinitiv	1				
(2) MSCI	**0.40**	1			
(3) S&P	**0.55**	**0.37**	1		
(4) ISS	**0.49**	**0.38**	**0.55**	1	
(5) Sustainalytics	**0.14**	**0.49**	0.07	0.15	1

Note: Correlations with significance levels <0.01 are in bold.

Table 4: Descriptive statistics before transformation of variables (Source: Own illustration)

Variable	N	Mean	Standard Deviation	Median	Skewness	Kurtosis
AF_DISP_0	3,968	49.38	342.99	0.43	13.12	226.66
ESG_Disagreement_0	3,968	1.97	0.86	1.98	0.08	2.46

bad, because it reduces the power of the statistical test or in other word the probability that the test will detect a difference that actually exists. Second, forecast dispersion is highly positively skewed. In this case, the tail region may act as an outlier for the statistical model. This is bad, because the outliers adversely affect the regression model's performance. Third, analyst forecast dispersion exhibits a high level of kurtosis. This means that the distribution of values is much more peaked than the normal and has heavy tails. This heavier tails leads to a few large outliers which are problematic for tests which rely on normality. As a result, differences are obscured, resulting in lower statistical power of the empirical test. To solve all these problems, the natural logarithm is used to normalize analyst forecast dispersion. As can be seen in Table 5 and Figure 6 in the annex, the normalized measure of analyst forecast dispersion is less skewed, less dispersed, has less kurtosis and resembles a bell-shaped normal distribution.

Table 5 presents descriptive statistics for the main sample, which consists of 3,968 firm-year observations. The dataset in Table 5 is not directly interpretable due to its log transformation of some variables. One way to obtain information about the central tendency and variability of the data set is to back-transform the data using the exponential function. The mean (median) of $ESG\_Disagreement_{i,t}$ is then 0.36 (0.43). Because $ESG\_Disagreement_{i,t}$ tends to be left-skewed, interpretating the variable using the median seems more appropriate to describe the central tendency. Because the variable ESG disagreement is itself a coefficient of variation, I find it difficult to interpret it using a descriptive statistics. The back-transformed standard deviation of $ESG\_Disagreement_{i,t}$, is 2.10. Therfore, with normal data, most of the observations are spread within one-fifth on each side of the mean.

7.2. Univariate Analysis

Table 6 shows the correlation coefficients. Correlation is a statistical measure that describes how two variables are related and indicates that as one variable changes in value, the other variable tends to change in a specific direction. The Pearson correlations are below and the Spearman coefficients above the line. Of particular interest for the empirical analysis are the Pearson coefficients. All variables in the empirical analysis exhibit a statistically significant linear correlation with $AF\_DISP_{i,t}$ at the 5% level. Thus, there appears to be a linear relationship between the variables which, based on the significance level, also applies to the population and not only to the sample. There is only a small chance that the results from the sample occurred due to chance (random sampling error). But this does not imply that there necessary is a cause and effect relationship. $ESG\_Disagreement_{i,t}$ and $AF\_DISP_{i,t}$ are weakly positivly correlated, as expected, with a value of 0.04. Thus, providing preliminary support that ESG disagreement is associated with an increase in analyst forecast dispersion.

Table 5: Descriptive statistics after transformation of variables (Source: Own illustration)

Variable	N	Mean	Standard Deviation	25%	Median	75%	Skewness	Kurtosis
AF_DISP	3,968	-2.51	1.11	-3.19	-2.59	-1.92	0.53	3.97
ESG_Disagreement	3,968	-1.01	0.74	-1.39	-0.84	-0.52	-0.90	3.55
Size	3,968	25.42	2.37	23.50	25.01	27.30	0.44	2.35
NANA	3,968	2.67	0.56	2.40	2.71	3.09	-0.83	4.10
BTM	3,968	0.0007	0.0008	0.0002	0.0004	0.0008	2.78	12.52
Earnings_VOL	3,968	5.65e+07	4.54e+08	2438233.9	1419371	8731285	5.87	38.95
Earnings_Surprise	3,968	37.20	3789.78	-5.76	0.66	11.15	-0.36	20.66
Leverage	3,968	0.25	0.17	0.12	0.25	0.37	0.36	2.45
ZMIJ	3,968	-3.16	1.10	-3.18	-3.18	-2.38	2.93	2.43
LOSS	3,968	0.087	0.28	0	0	0		

Table 6: Pearson/Spearman Correlation Coefficients (Source: Own illustration)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) AF_DISP	1	**0.04**	**0.03**	**-0.05**	**0.28**	**0.17**	**-0.12**	**0.13**	**0.20**	**0.33**
(2) ESG_Disagreement	**0.04**	1	-0.01	**-0.18**	**-0.05**	-0.03	**0.03**	**-0.07**	**-0.07**	0.01
(3) Size	**0.05**	**-0.03**	1	**0.14**	-0.003	**0.87**	0.01	**-0.19**	**-0.19**	**-0.10**
(4) NANA	**-0.07**	**-0.18**	**0.16**	1	**-0.13**	**0.11**	0.01	**-0.07**	**-0.08**	0.003
(5) BTM	**0.23**	**0.08**	**-0.05**	**-0.07**	1	**0.32**	-0.03	**0.10**	**0.18**	**0.09**
(6) Earnings_VOL	**0.08**	-0.0005	**0.48**	**0.11**	**0.11**	1	0.007	**-0.11**	**-0.09**	-0.02
(7) Earnings_Surprise	**-0.08**	**0.04**	0.01	0.01	**-0.04**	**0.04**	1	-0.03	**-0.10**	**-0.28**
(8) Leverage	**0.13**	**-0.06**	**-0.19**	**-0.08**	**0.09**	0.01	-0.02	1	**0.97**	**0.13**
(9) ZMIJ	**0.21**	**-0.06**	**-0.19**	**-0.10**	**0.15**	-0.001	**-0.07**	**0.97**	1	**0.25**
(10) LOSS	**0.39**	0.006	**-0.10**	-0.005	**0.09**	**0.04**	**-0.22**	**0.14**	**0.26**	1

Note: Pearson (Spearman) coefficients are below (above). Correlations with significance levels <0.05 are in bold.

Furthermore, there is a very strong positive correlation between Leverage and ZMIJ with a value of 0.97 at the 5% significance level. This correlation exists because part of the calculation of the Zmijewski financial distress score includes a debt-to-assets ratio. The question is whether this correlation will lead to problems for the recession model due to multicollinearity. Multicollinearity occurs when independent variables in a regression model are correlated. This correlation is a problem because independent variables should be independent. If the degree of correlation between the two variables is high enough, it can cause problems when fitting the model and interpreting the results because it affects the coefficients and p-values. Therefore, it must be decided whether one of the two variables should be excluded or both should be kept in the model. However, multicollinearity only affects the specific variables that are correlated. A model can have severe multicollinearity and yet some variables in the model can be completely unaffected. Therefore if multicollinearity exists for the control variables but not the experimental variables, the experimental variables can be interpreted without problems. Because the multicollinearity is not present in $ESG\_Disagreement_{i,t}$, the variable of interest, the issue of multicollinearity does not need to be resolved and both variables are kept in the sample. AF_DISP and LOSS also show a moderate positive correlation with a value of 0.39 at the 5% significance level. But, there is no risk of multicollinearity as ESG_Disagreement and LOSS show almost no correlation with a value of 0.006. Nor is the correlation significant at the 5% significance level.

In some cases, however, the Pearson coefficient is not appropriate. In this case, the Spearman coefficient is used. The Spearman's rank correlation is used when Pearson's correlation cannot be run due to violations of normality, a non-linear relationship or when ordinal variables are being used. To analyze the data using the Spearman coefficient, two assumptions must be met. Otherwise, the Spearman correlation may not produce valid results. First, the two variables should be measured on an ordinal or continuous scale. Second, there needs to be a monotonic relationship between the two variables. In this empirical study, Earninngs_Vol and BTM have a non-normal distribution. Therefore, the Pearson coefficient is not appropriate and the Spearman coefficient is used instead. Earnings_VOL and Size show a very strong positive correlation with a value of 0.87, which is significant at the 5% level. BTM and Earnings_VOL exhibit a weak positive correlation with a value of 0.32, which is also significant at the 5% level. In both cases, multicollinearity is not an issue.

7.3. Multivariate Analyses

Table 7 reports the OLS regression results. To test the goodness of fit of the regression model, R-squared is used. The R-squared value indicates how well the model explains the dependent variable's variance. The first model has a R-squared value of 0.001. Therefore, the model does not produce predictions that are reasonably precise. But, because the predictor of ESG_Disagreement is statistically significant, it can still be concluded that changes in ESG_Disagreement are associated with changes in AF_DISP. One limitation of R-squared is that it is invalid for nonlinear regression. To test whether the regression is linear, I examine the residuals plot (see Figure 7 in the annex). The residuals plot shows no signs of nonlinearity. I also examine the Significance F of the overall regression model. The Significance F represents the p-value for the overall regression model. This test shows whether a model with all its independent variables explains the variability of the dependent variable better than a model without any independent variables. The Significance of F for the first model is 0.019. Because 0.019 is lower than the 5% significance level, the regression model as a whole is statistically significant, i.e. the model fits the data better than the model with no predictor variables. The coefficient of ESG_Disagreement is 0.0557. Thus, the results suggest that ESG_Disagreement is positively associated with AF_Dispersion, supporting the hypothesis. Because both variables were transformed with the natural logarithm, changes cannot be expressed in absolute numbers, but only as percentages. Hence, the results suggest that when ESG disagreement increases by 1%, analyst forecast dispersion also increases by 5.57%.

The second model controls for the influence of other factors on analysts' forecast dispersion. I use only three control variables to address the problem of underspecification of the first model. An underspecified model, i.e., a model that is too simple, can lead to biased estimates. The second model also avoids the problem of overspecification caused by too many variables. A model that contains to many variables, i.e. is too complex, tends to reduce the precision of coefficient estimates and predicted values. Furthermore, I use the adjusted R-square to test the goodness of fit of the regression model. The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. It takes into account whether R-squared is higher because the predictors are better or just because the model has more predictors. The second model has an adjusted R-squared of 0.091. It is therefore compared to the first model better at explaining the variability in analyst forecast dispersion. The coefficient of ESG_Disagreement is 0.0534. The association between ESG disagreement and analyst forecast dispersion remains positive and significant at the 5% level. The third model adds additional control variables. It has an adjusted R-squared value of 0.223. Figures 8 and 9 in the annex show the residual plots for models two and three, respectively. Both residual plots show no sign of a nonlinear regression. The coefficient of ESG_Disagreement is 0.0408 and is statistically significant at the 10% level. To control for heteroskedasticity, I include robust standard errors in all three models. Heteroscedasticity refers to the unequal scatter of residuals. When heteroscedasticity is present in a regression analysis, the results of the analysis become hard to trust. Heteroscedasticity increases the variance of the regression coefficient estimates, but the regression model does not take this into account. As a result, a regression model is more likely to declare a parameter in the model to be statistically significant when it is in fact not. The fourth model controls for year-fixed and country-fixed effects. Fixed effects are commonly used in panel data analysis, to account for unobserved heterogeneity. By including fixed effects, individual-specific factors (e.g. individual attitudes, preferences and abilities) that remain constant over time are controlled for, ensuring that the regression results are not distorted by these unobserved factors. Fixed effects models thus help to mitigate the problem of omitted variable bias, which arises when important variables are excluded from the analysis (Collischon & Eberl, 2020, p. 291 f.). The fourth model has an adjusted R-squared value of 0.375. The coefficient of ESG_Disagreement is 0.0131. However, the coefficient is no longer statistically significant at the 10% level. Therefore, after removing the time-constant error term, there is no longer a significant positive relationship between ESG disagreement and analyst forecast dispersion. Hence, after accounting for unobserved heterogeneity, it is not possible to reach a conclusion about whether ESG disagreement is associated with the dispersion of analysts' forecasts.

8. Robustness Checks

One concern is the impact of outliers on the regression model. As discussed in Chapter 7.1, outliers or those treated as outliers (i.e., fat tails) can reduce the power of the regression model. In my regression model, there are several control variables that might affect its performance. First, the book-to-market-ratio is skewed. Second, earnings volatility and earnings surprise exhibit outliers (see Table 5). To test whether these control variables reduce the statistical power of the regression model, I transform all three variables with the natural logarithm and repeat the main analysis (see Table 22 in the annex). The first model uses only size, ln(BTM), and ZMIJ as control variables. It has an adjusted R-squared value of 0.091. The coefficient for ESG_Disagreement is 0.084. A 1% increase in ESG disagreement is associated with an 8.47% increase in analyst forecast dispersion. The coefficient is statistically significant at the 0.1% level, making it very unlikely that the result is due to chance. Next, I use all control variables. The second model has an adjusted R-squared value of 0.199. The coefficient of ESG_Disagreement is 0.042. The result is not statistically significant at the 10% level anymore. The third model again includes year-fixed and industry-fixed effects. It has an adjusted R-squared value of 0.379. The coefficient for ESG_Disagreement is 0.019. However, the result is also not statistically significant at the 10% level. The results are consistent with the main analysis.

Table 7: Regression results of the main analysis (Source: Own illustration)

	(1) AF_DISP	(2) AF_DISP	(3) AF_DISP	(4) AF_DISP
ESG_Disagreement	0.0557**	0.0534**	0.0408*	0.0131
	(0.019)	(0.020)	(0.060)	(0.540)
Size		0.0387***	0.0450***	-0.106***
		(0.000)	(0.000)	(0.000)
NANA			-0.086***	0.0804**
			(0.006)	(0.015)
BTM		305.9***	225.3***	231.4***
		(0.000)	(0.000)	(0.000)
Earnings_VOL			05.64e-11	3.47e-10**
			(0.563)	(0.036)
Earnings_Surprise			0.00014*	0.00012*
			(0.067)	(0.054)
Leverage			-3.717***	-4.064***
			(0.000)	(0.000)
ZMIJ		0.197***	0.684***	0.729***
		(0.000)	(0.000)	(0.000)
LOSS			1.085***	0.926***
			(0.000)	(0.000)
Year-Fixed Effects	No	No	No	Yes
Country-Fixed Effects	No	No	No	Yes
N	3,968	3,968	3,968	3,968
R-Square	0.001	0.092	0.225	0.385
Adjusted R-Square	0.001	0.091	0.223	0.375
F-Statistic	0.019	0.000	.	0.000

Note: P-values are below the coefficients in brackets. The significance levels are market with stars: * p<0.10, ** p<0.05, *** p<0.01.

Another concern is the influence of financial firms and utilities on the empirical results. It is a common approach in empirical finance to exclude financial firms. This is because their business model is highly different from other firms. Fama and French (1992, p. 429) state: "We exclude financial firms because the high leverage that is normal for these firms probably does not have the same meaning as for non-financial firms, where high leverage more likely indicates distress." Utilities are excluded due to their association with the state. State-owned firms are often not profit-oriented and are highly affected by governmental decisions. Their business model also differs from that of private firms in that they perform public functions. Utilities also have a very high leverage and an unusually high book-to-market ratios, which makes them highly sensitive to interest rate changes. For these reasons, utilities are usually excluded from empirical studies (Stack Exchange, 2023). Table 23 in the annex reports the empirical results without financials. The first model without control variables has an adjusted R-squared value of 0.001. The coefficient for ESG_Disagreement is 0.0606. The coefficient is statistically significant at the 5% level with a value of 0.010. The second model with three control variables has an adjusted R-squared value of 0.090. The coefficient for ESG_Disagreement is 0.0578. The coefficient is statistically significant at the 5% level with a value of 0.011. The third model with control variables has an adjusted R-squared value of 0.224. The coefficient for ESG_Disagreement is 0.0455. The coefficient is also statistically significant at the 5% level with a value of 0.034. The fourth model with fixed-effects has an adjusted R-squared value of 0.376. The coefficient for ESG_Disagreement is 0.0198. The coefficient is not statistically significant at the 10% level. Table 24 reports the empirical results without financial, utilities and real estate firms. Real estate firms are excluded due to their unusual high leverage. The first model has an adjusted R-squared value of 0.001. The coefficient for ESG_Disagreement is 0.0623. The coefficient is statistically significant at the 5% level with a value of 0.012. The second model has an adjusted R-squared value of 0.103. The coefficient for ESG_Disagreement is 0.0675. The coefficient is statistically significant at the 1% level with a value of 0.004. The third model has an adjusted R-squared value of 0.235. The coefficient for ESG_Disagreement is 0.0507. The coefficient is statistically significant at the 5% level with a value of 0.024. The fourth model has an adjusted R-squared value of 0.383. The coefficient for ESG_Disagreement is 0.0179. The coefficient is not statistically significant at the 10% level. The results of the main analysis are thus robust to the influence

of financial firms, utilities, and real estate firms.

Afterwards, I test whether the relationship between ESG disagreement and analyst forecast dispersion is consistent over time. Table 8 shows the results. In 2018, the coefficient of ESG_Disagreement is 0.143. The coefficient is highly statistically significant at the 0.1% level. As a result, it is unlikely that the result is due to chance. A 1% increase in ESG disagreement leads to a 14.3% increase in analyst forecast dispersion. In 2019, the coefficient of ESG_Disagreement is 0.0938. The coefficient remains statistically significant at the 5% level with a value of 0.016. In 2020, the coefficient of ESG_Disagreement is 0.0281. However, the results are not significant at the 10% level. Accordingly, it cannot be ruled out that the results are due to chance. In 2021, the relationship between ESG_Disagreement and AF_DISP turns negative with a value of -0.02263. The result is also not statistically significant at the 10% level. In 2022, there is an even stronger negative association between ESG_Disagreement and AF_DISP with a coefficient of -0.0930. The results remain not statistically significant at the 10% level. When examining the coefficients, it is evident that the relationship between ESG_Disagreement and AF_DISP is at first positive, but becomes negative over time. Also, with the exception of 2022, the strength of the relationship decreases over time. Moreover, the relationship between ESG_Disagreement AF_DISP is highly statistically significant in the years 2018 and 2019, but not significant in the last three years. Thus, the relationship between ESG disagreement and analyst forecast dispersion does not appear to be consistent over time. Next, I incorporate fixed effects and repeat the regression model. Table 25 in the appendix shows the results. Across all years, the coefficient for ESG disagreement remains statistically insignificant. The results remain robust to the main analysis.

Another concern is whether the association between ESG disagreement and the dispersion of analysts' forecasts is robust for different measures of ESG Disagreements. For this reason, I use three additional measures of ESG Disagreement that differ from ESG_Disagreement in their calculation (see Table 9). ESG_Disagreement_3 includes three ESG ratings, while ESG_Disagreement_4 and ESG_Disagreement_5 include four and five ESG ratings, respectively. In the first model without control variables, the coefficient for ESG_disagreement_3 is 0.0517 and significant at the 5% level. The second model with three control variables has a coefficient of 0.0440 and is significant at the 10% level. The third model with all control variables, ESG_Disagreement has a coefficient of 0.0349. The result is not significant at the 10% level. The fourth model with year fixed and country fixed effects has a coefficient for ESG_Disagreement_3 of 0.0180. It is not statistically significant at the 10% level with a value of 0.391, so there is a high chance that the result is due to chance.

The results for ESG_disagreement_4 and ESG_disagreement_5 are both highly not statistically significant. Thus, there is a high probability that the results are due to chance. The results with control variables are tabulated in model five and six in Table 9. As can be seen, the number of observations for both models is rather low. A larger sample may provide more precise estimates and more significant results.

Next, I disaggregate ESG ratings into its pillar scores to examine the extent to which environmental, social, and governmental issues account for the influence of ESG disagreement on analyst forecast dispersion. As can be seen in Table 10, all results for the influence of the pillar scores are highly non-significant. Thus, there is a high probability that the results are due to chance. Consequently, no conclusive statement can be made about which issues are driving the positive association between ESG_Disagreement and analyst forecast dispersion.

In total, the findings of this master thesis do not provide sufficient evidence to support a significant association between ESG disagreement and analyst forecast dispersion. Thus, it cannot be concluded with confidence that the disagreement among ESG rating agencies influences the dispersion of analyst EPS forecasts. In other words, the disagreement between ESG rating agencies regarding a firm's non-financial ESG performance does not seem to have a discernible impact on analysts' uncertainty about the firm's future earnings. These results have important implications for practitioners, suggesting that non-financial ESG criteria may not play a substantial role in analysts' evaluation of a firm's financial performance.

9. Limitations and Future Research Opportunities

9.1. Limitations

There are several limitations that may affect the validity and reliability of the results of this empirical study. One limitation is that the results are biased by the selection of ESG rating providers. In this empirical study, ESG ratings from Refinitiv Eikon, MSCI, S&P, ISS and Sustainalytics were used. The results may not be reproducible with ESG ratings from other ESG rating providers, as the degree of disagreement between ESG rating providers varies. Another limitation is the omitted variable bias caused by confounding variables. Omitted variable bias refers to the bias that can occur in regression analysis when an important independent variable is left out of the model. The omitted variable bias occurs because cofounding variables are still affecting the dependent variable, but their effects are absorbed by the error term in the regression model. For example, analyst forecast dispersion could be influenced by the forecast horizon. The further away a forecast is from a firm's actual earnings announcement, the more uncertain the forecast (Chopra, 1998, p. 37). Although the inclusion of fixed effects helps to mitigate endogeneity concerns and control for time-constant factors, it can not completely eliminate the possibility of omitted variable bias. Fixed effects models assume that the unobserved heterogeneity across firms and time periods is adequately captured by the fixed effects variables. However, if there are additional unobserved variables that are correlated with both

Table 8: ESG Disagreement over time (Source: Own illustration)

	(1) AF_DISP -2018-	(2) AF_DISP -2019-	(3) AF_DISP -2020-	(4) AF_DISP -2021-	(5) AF_DISP -2022-
ESG_Disagreement	0.143***	0.0938**	0.0281	-0.0263	-0.0930
	(0.000)	(0.016)	(0.533)	(0.569)	(0.441)
Size	0.0242	0.0534***	0.0909***	-0.0216	0.00126
	(0.117)	(0.000)	(0.000)	(0.169)	(0.957)
NANA	-0.0420	0.0606	-0.204***	-0.126	0.0465
	(0.524)	(0.338)	(0.002)	(0.169)	(0.602)
BTM	245.8***	208.6***	213.5***	159.4***	237.3**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.012)
Earnings_VOL	3.75e-10*	1.79e-11	-3.55e-10**	4.11e-10*	-2.28e-10
	(0.067)	(0.924)	(0.040)	(0.072)	(0.742)
Earnings_Surprise	0.00005	-0.0002	0.00003	0.0005***	0.0005***
	(0.760)	(0.145)	(0.841)	(0.000)	(0.006)
Leverage	-6.121***	-6.641***	-3.976***	-1.691*	-0.602
	(0.000)	(0.000)	(0.000)	(0.055)	(0.650)
ZMIJ	1.051***	1.148***	0.792***	0.309**	0.133
	(0.000)	(0.000)	(0.000)	(0.026)	(0.520)
LOSS	0.665***	0.971	0.982***	1.224***	1.591***
	(0.001)	(0.262)	(0.000)	(0.000)	(0.000)
N	785	1,048	1,053	752	330
R-Square	0.228	0.213	0.277	0.232	0.257
Adjusted R-Square	0.219	0.206	0.270	0.222	0.236

Note: P-values are below the coefficients in brackets. The significance levels are market with stars: * p<0.10, ** p<0.05, *** p<0.01.

ESG disagreement and analyst forecast dispersion, the estimated coefficients may still be biased by time-varying heterogeneity (Collischon & Eberl, 2020, p. 292 f.). There is also the problem of endogeneity. Endogeneity refers to the situation in which the independent variables in a regression model are affected by the dependent variable. This can create a circular relationship between the variables, causing a bias in the estimates of the model parameters. Financial analysts use financial and non-financial information to evaluate a firm's financial performance. ESG rating providers may use financial analysts' non-financial assessment as an input for their own ESG performance assessment. Another problem is the opaqueness of ESG rating agencies' methodologies. ESG rating agencies attach different weights to their pillar scores. This makes it difficult to compare the pillar scores of different ESG rating providers if they are not directly comparable, as is the case with Sustainalytics. The assumption that the pillar scores are equally weighted distorts the results if the true weightings vary. Another problem of this empirical study is the data availability for the E/S/G pillar scores. The small sample size resulting from the lack of freely available pillar scores leads to inaccurate conclusions and reduces the generalizability of the results.

9.2. Future Research Opportunities

Future researchers could address some of the limitations mentioned in the previous chapter. They could address the selection bias by using different ESG rating providers in their study to see if the results remain consistent as the level of disagreement between ESG rating providers changes. They could also use different control variables in their empirical study that could better reflect changes in the dispersion of analyst forecasts and thus reduce the risk of variable omission. They further could address endogeneity concerns by controlling for prior disclosed nonfinancial information from financial analysts. In addition, future researchers could use pillar scores that are directly comparable or find a way to make pillar scores more comparable. To some extent, improved transparency in ESG rating providers' methodologies should also help. Or, they could examine whether environmental, social, or governance issues are responsible for the positive association between ESG disagreement and analyst forecast dispersion. Another interesting research direction would be to examine how changes in ESG disagreement affect the dispersion of analysts' forecasts. In this study, the levels of ESG disagreement were analyzed. However, certain information from the previous ESG ratings may already available be to financial analyst. Therefore, changes in ESG information in particular may be responsible for variations in analysts' forecasts. Yet another interesting research direction would be to

Table 9: Alternative measures of ESG disagreement (Source: Own illustration)

	(1) AF_DISP	(2) AF_DISP	(3) AF_DISP	(4) AF_DISP	(5) AF_DISP	(6) AF_DISP
ESG_Disagreement_3	0.0517**	0.0440*	0.0349	0.0180		
	(0.029)	(0.053)	(0.107)	(0.391)		
ESG_Disagreement_4					-0.0227	
					(0.823)	
ESG_Disagreement_5						-0.0548
						(0.799)
Size		0.0458***	0.0515***	-0.106***	-0.0641	-0.0805
		(0.000)	(0.000)	(0.000)	(0.446)	(0.385)
NANA			-0.0832**	0.0790**	0.139	0.107
			(0.012)	(0.021)	(0.415)	(0.642)
BTM		307.6***	223.4***	232.0***	292.1*	251.8
		(0.000)	(0.000)	(0.000)	(0.105)	(0.213)
Earnings_VOL			1.14e-11	3.34e-10**	2.35e-10	3.52e-11
			(0.906)	(0.041)	(0.856)	(0.981)
Earnings_Surprise			1.27e-4	1.09e-4*	7.16e-4*	6.60e-4
			(0.111)	(0.085)	(0.063)	(0.138)
Leverage			-4.078***	-4.495***	-1.987	-2.247
			(0.000)	(0.000)	(0.256)	(0.268)
ZMIJ		0.202***	0.746***	0.799***	0.445*	0.468
		(0.000)	(0.000)	(0.000)	(0.100)	(0.138)
LOSS			1.054***	0.882***	1.155***	1.114***
			(0.000)	(0.000)	(0.000)	(0.001)
Year-Fixed Effects	No	No	No	Yes	Yes	Yes
Country Fixed Effects	No	No	No	Yes	Yes	Yes
N	3,783	3,783	3,783	3,783	141	127
R-Square	0.001	0.096	0.228	0.392	0.428	0.405
Adjusted R-Square	0.001	0.095	0.227	0.382	0.344	0.303

Note: P-values are below the coefficients in brackets. The significance levels are market with stars: * p<0.10, ** p<0.05, *** p<0.01.

investigate the influence of non-financial disclosure regulations on the association between ESG disagreement and analyst forecast dispersion. Researchers could use a difference-in-difference design to control whether the introduction of a non-disclosure regulation is associated with greater analyst forecast dispersion. There are two non-financial disclosure regulations that are of particular interest. On is the European Union's NFRD, which requires all firms covered by the directive to report for the first time for the 2017 financial year on non-financial issues (Hankamper-Vandebulcke, 2021, p. 4). The other is an amendment to the Financial Instruments Exchange Act of Japan, which requires listed firms in Japan with a current fiscal year-end to report on ESG issues by March 2023 (Tomoko & Kyoko, 2022). Both are of interest for a difference-in-difference design. Unfortunately, due to the chosen time period of this sample, it is not possible to apply such a difference-in-difference design to this empirical study.

10. Conclusion

Non-financial ESG information has become an increasingly important source of information for the investment community, as it allows for a more thorough assessment of a firm's long-term risks and opportunities. One important group that relies on non-financial ESG information are financial analysts. Financial analysts use non-financial information alongside traditional financial information to inform their forecasts. However, ESG information often lack standardization and are difficult to compare. For this reason, financial analysts increasingly rely on ESG rating agencies as third-party information intermediaries to make sense of available ESG information. ESG rating agencies aggregate the available ESG information and produce ESG ratings by assessing a firm's ESG performance. Those ESG ratings intend to inform investors about a firm's ability to cope with long-term risks and opportunities. However, ESG rating agencies disagree on what constitutes as good ESG performance. This leads to sometimes widely divergent ESG ratings. The reason ESG rating agencies tend to disagree is

Table 10: E/S/G Disagreement (Source: Own illustration)

	(1) AF_DISP	(2) AF_DISP	(3) AF_DISP	(4) DISP	(5) DISP	(6) AF_DISP
E_Disagreement	-0.0340	0.0224				
	(0.734)	(0.839)				
S_Disagreement			-0.0813	-0.0308		
			(0.536)	(0.777)		
G_Disagreement					-0.0893	-0.0206
					(0.465)	(0.862)
Size	0.0436	-0.00984	0.0360	-0.0187	0.0336	-0.0137
	(0.441)	(0.931)	(0.541)	(0.872)	(0.552)	(0.904)
NANA	-0.277	-0.381	-0.282	-0.403	-0.281	-0.403
	(0.441)	(0.343)	(0.248)	(0.306)	(0.250)	(0.306)
BTM	467.1***	386.8	423.8**	351.9	438.5***	367.1
	(0.001)	(0.105)	(0.014)	(0.165)	(0.004)	(0.133)
Earnings_VOL	-6.58e-10	-3.84e-10	-4.29e-10	-2.65e-10	-4.54e-10	-3.05e-10
	(0.319)	(0.796)	(0.541)	(0.862)	(0.525)	(0.842)
Earnings_Surprise	0.0004**	0.0003	0.0004**	0.0003	0.0004**	0.0003
	(0.024)	(0.155)	(0.019)	(0.151)	(0.029)	(0.157)
Leverage	-2.142	-1.314	-2.356	-1.390	-2.238	-1.350
	(0.470)	(0.585)	(0.432)	(0.565)	(0.444)	(0.575)
ZMIJ	0.322	0.328	0.354	0.329	0.327	0.322
	(0.479)	(0.389)	(0.443)	(0.387)	(0.465)	(0.398)
LOSS	0.926***	0.968*	0.929***	0.984*	0.910***	0.977**
	(0.006)	(0.093)	(0.004)	(0.088)	(0.008)	(0.090)
Year-Fixed Effects	No	Yes	No	Yes	No	Yes
Country Fixed Effects	No	Yes	No	Yes	No	Yes
N	83	79	83	79	83	79
R-Square	0.236	0.413	0.241	0.414	0.241	0.413
Adjusted R-Square	0.142	0.262	0.148	0.262	0.148	0.262

Note: P-values are below the coefficients in brackets. The significance levels are market with stars: * p<0.10, ** p<0.05, *** p<0.01.

because they differ in scope, weighting, and measurement of ESG information. Because ESG rating agencies compete with each other for market share, there is no single approach to ESG rating methodologies. In addition, ESG rating methodologies are not fully transparent. As financial analysts seek to understand ESG ratings and their underlying data, they obtain their own private information, leading to divergent opinions about a firm's long-term risks and opportunities.

The objective of this master thesis was to empirically investigate the influence of ESG rating disagreement on analyst forecast dispersion in an international setting. Prior research based on Kimbrough et al. (2022) found a positive association between ESG rating disagreement and analyst forecast dispersion for firms in the US. The first regression model without control variables shows that there is indeed a statistically significant relationship between ESG disagreement and analyst forecast dispersion. The coefficient of ESG disagreement is 0.0557 and is statistically significant at the 5% level. Because both ESG disagreement and analyst forecast dispersion were transformed with the natural logarithm, a 1% increase in ESG disagreement is associated with a 5.57% increase in analyst forecast dispersion. However, the first model has a low R-squared value and therefore does not produce predictions that are reasonably precise. The introduction of control variables increases the predictability of the empirical model. The second (third) model with three (eight) control variables are also statistically significant at the 5% (10%) level. A 1% increase in ESG disagreement is associated with a 5,34% (4,08%) increase in analyst forecast dispersion. However, the inclusion of year and country fixed effects within the regression model leads to a notable shift in the nature of the obtained results, yielding statistically non-significant findings. To ensure the validity and reliability of these findings, I employ several robustness checks. First, I address the presence of skewed distributions in some of the control variables by applying a natural logarithm transformation. This transformation helps to control for outliers that might influence the regression model. After implementing this adjustment, the results remain consistent with the main findings, providing additional confidence in the robustness of the findings. Second, I exclude financial firms and utilities from the analysis due to their fundamentally distinct na-

ture from private firms. Additionally, real estate firms are excluded due to their unusually high levels of leverage. Despite these exclusions, the results remain consistent with the main findings, reinforcing the stability of the observed relationships. Third, I examine the time consistency of the relationship between ESG disagreement and analyst forecast dispersion. However, there is a deviation from the main results in the years 2018 and 2019. This incon-sistency prompts further investigation into the potential factors driving the variation and underscores the need for cautious interpretation of the more distant results. Fourth, alternative measures of ESG disagreement are employed to assess their impact on the results. Despite these variations in measurement, the main findings remain unchanged, indicating robustness in the relationship between ESG disagreement and analyst forecast dispersion. Fifth, I explore the individual influence of environmental, social, and governance factors on analyst forecast dispersion. However, due to the small sample size, the results are not statistically significant and cannot be considered representative. Overall, the empirical results remain robust after performing several robustness checks. Hence, no definitive conclusion can be drawn regarding the influence of ESG disagreement on the dispersion of analysts' forecasts.

These findings hold significant implications for practitioners, particularly those involved in the investment industry, as they challenge the relevance of non-financial ESG information provided by ESG rating agencies in informing financial analysts' forecasts. This master thesis also presents opportunities for further research in the field. Potential avenues include investigating the influence of environmental, social, and governance (ESG) criteria on analyst forecast dispersion, or employing a difference-in-difference design to study the effects of new non-financial disclosure requirements. For instance, researchers could explore the impact of regulatory frameworks like the European Non-Financial Reporting Directive (NFRD), which predates the sample period covered in this study, or the recent amendment to the Financial Instruments Exchange Act of Japan which mandates listed firms in Japan to include ESG information in their current fiscal year reporting by March 2023.

References

Abarbanell, J. S., Lanen, W. N., & Verrecchia, R. E. (1995). Analysts' forecasts as proxies for investor beliefs in empirical research. *Journal of Accounting and Economics*, *20*(1), 31–60.

Ackert, L. F. (1997). Prior Uncertainty, Analyst Bias, and Subsequent Abnormal Returns. *The Journal of Financial Research*, *20*(2), 263–273.

Ajinkya, B. B., Atiase, R. K., & Gift, M. J. (1991). Volume of Trading and the Dispersion in Financial Analysts' Earnings Forecasts. *The Accounting Review*, *66*(2), 389–401.

Allen, K. (2018). Lies, damned lies and ESG rating methodologies. Financial Times. Retrieved September 30, 2022, from https://www.ft.com/content/2e49171b-a018-3c3b-b66b-81fd7a170ab5

Asquith, P., & Weiss, L. A. (2016). *Lessons in Corporate Finance*. New Jersey.

Athanassakos, G., & Kalimipalli, M. (2003). Analyst Forecast Dispersion and Future Stock Return Volatility. *Quarterly Journal of Business and Economics*, *42*(1 & 2), 57–78.

Avramov, D., Chordia, T., Jostova, G., & Philipov, A. (2009). Dispersion in analysts' earnings forecasts and credit rating. *Journal of Financial Economics*, *91*(1), 83–101.

Barefield, R. M., & Comiskey, E. E. (1975). The Accuracy of Analysts' Forecasts of Earnings per Share. *Journal of Business Research*, *3*(3), 241–252.

Barker, R. G. (1998). The market for information—evidence from finance directors, analysts and fund managers. *Accounting and Business Research*, *29*(1), 3–20.

Barron, O. E., Kim, O., Lim, S. C., & Douglas, E. S. (1998). Using Analysts' Forecasts to Measure Properties of Analysts' Information Environment. *The Accounting Review*, *73*(4), 421–433.

Barron, O. E., Standford, M. H., & Yu, Y. (2010). Further Evidence on the Relation between Analysts' Forecast Dispersion and Stock Returns. *Contemporary Accounting Research*, *26*(2), 329–357.

Barry, C. B., & Jennings, R. H. (1992). Information and Diversity of Analyst Opinion. *The Journal of Financial and Quantitative Analysis*, *27*(2), 169–183.

Beckers, S., Steliaros, M., & Thomson, A. (2004). Bias in European Analysts' Earnings Forecasts. *Financial Analysts Journal*, *60*(2), 74–85.

Behn, B. K., Choi, J.-H., & Kang, T. (2008). Audit Quality and Properties of Analyst Earnings Forecasts. *The Accounting Review*, *83*(2), 327–349.

Belnap, A. (2022). The effect of intermediary coverage on disclosure: Evidence from a randomized field experiment [Corrected proof in press]. *Journal of Accounting and Economics*.

Benston, G. J. (1973). AssociationRequired Disclosure and the Stock Market: An Evaluation of the Securities Exchange Act of 1934. *The American Economic Review*, *63*(1), 132–155.

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

Berk, J., & van Binsbergen, J. H. (2021). *The Impact of Impact Investing* [Working Paper No. 3981, Stanford University Graduate School of Business].

Beunza, D., & Garud, R. (2007). Calculators, Lemmings or Frame-Makers? The Intermediary Role of Securities Analysts. *The Sociological Review*, *55*(2), 13–39.

Bloomberg. (2022). ESG May Surpass $41 Trillion Assets in 2022, But Not Without Challenges, Finds Bloomberg Intelligence. Press Release. Retrieved December 6, 2022, from https://www.bloomberg.com/company/press/esg-may-surpass-41-trillion-assets-in-2022-but-not-without-challenges-finds-bloomberg-intelligence/

BNP Paribas. (2019). The ESG Global Survey 2019. Retrieved December 5, 2022, from https://securities.cib.bnpparibas/app/uploads/sites/3/2019/08/esg-global-survey-en-2019.pdf

Boffo, R., & Patalano, R. (2020). ESG Investing: Practices, Progress and Challenges. Retrieved December 5, 2022, from www.oecd.org/finance/ESG-Investing-Practices-Progress-and-Challenges.pdf

Brackley, A., Brock, E. K., & Nelson, J. (2022). Rating the Raters Yet Again: Increasing ESG Scrutiny Makes Current Rate the Raters Study Even More Crucial. Sustainability. Retrieved December 15, 2022, from https://www.sustainability.com/thinking/rating-the-raters-yet-again-increasing-esg-scrutiny-makes-current-rate-the-raters-study-even-more-crucial/

Brown, R., Chan, H. W. H., & Ho, Y. K. (2009). Analysts' recommendations: from which signal does the market take its lead? *Review of Quantitative Finance and Accounting*, *33*, 91–111.

Byard, D., Li, Y., & Yu, Y. (2011). The Effect of Mandatory IFRS Adoption on Financial Analysts' Information Environment. *Journal of Accounting Research*, *49*(1), 69–96.

Capital Group. (2021). ESG Global Study 2021. Retrieved January 12, 2023, from https://www.capitalgroup.com/content/dam/cgc/tenants/eacg/esg/global-study/esg-global-study-full-report(en).pdf

Capstaff, J., Paudyal, K., & Rees, W. (1995). The Accuracy and Rationality of Earnings Forecasts by UK Analysts. *Journal of Business Finance & Accounting*, *22*(1), 67–85.

CFA Institute. (2020). Future of Sustainability in Investment Management: From Ideas to Reality. Retrieved November 5, 2022, from https://www.cfainstitute.org/-/media/documents/survey/future-of-sustainability.pdf

Chatterji, A., & Levine, D. (2006). Breaking Down the Wall of Codes: Evaluating Non-Financial Performance Measurement. *California Management Review*, *48*(2), 29–51.

Cho, S.-Y., Lee, C., & Pfeiffer, R. J. (2013). Corporate Social Responsibility Performance Information and Information Asymmetry. *Journal of Accounting and Public Policy*, *32*(1), 71–83.

Chopra, V. K. (1998). Why So Much Error in Analysts' Earnings Forecasts? *Financial Analysts Journal*, *54*(6), 35–42.

Christensen, D., Serafeim, G., & Sikochi, A. (2021). Why is Corporate Virtue in the Eye of The Beholder? The Case of ESG Ratings. *Accounting Review*, *97*(1), 147–175.

Cision. (2020). Morningstar to Acquire Sustainalytics and Expand Access to ESG Research, Data, and Analytics for Investors Worldwide. PR Newswire. Retrieved November 25, 2022, from https://www.prnewswire.com/news-releases/morningstar-to-acquire-sustainalytics-and-expand-access-to-esg-research-data-and-analytics-for-investors-worldwide-301044196.html

Collischon, M., & Eberl, A. (2020). Let's Talk About Fixed Effects: Let's Talk About All the Good Things and the Bad Things. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *72*(2), 290–299.

Cornell, B. (2020). ESG preferences, risk and return. *European Financial Management*, *27*(1), 12–19.

Cowles, A. (1933). Can Stock Market Forecasters Forecast? *Econometrica*, *1*(3), 309–324.

Cui, J., Jo, H., & Na, H. (2018). Does Corporate Social Responsibility Affect Information Asymmetry? *Journal of Business Ethics*, *148*(3), 549–572.

D. E. Shaw. (2022). Keep the Change: Analyzing the Increase in ESG Ratings for U.S. Equities. Retrieved January 3, 2023, from https://www.deshaw.com/assets/articles/DESCO_Market_Insights_ESG_Ratings_20220408.pdf

De Bondt, W. F. M., & Forbes, W. P. (1999). Herding in analyst earnings forecasts: evidence from the United Kingdom. *European Financial Management*, *5*(2), 143–163.

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1996). Causes and Consequences of Earnings Manipulation: An Analysis of Firms Subject to Enforcement Actions by the SEC. *Contemporary Accounting Research*, *13*(1), 1–36.

Deloitte. (2021). ESG Ratings: do they add value? How to get prepared? Deloitte. Retrieved November 28, 2022, from https://www2.deloitte.com/ce/en/pages/about-deloitte/articles/esg-ratings-do-they-add-value.html

Deloitte. (2022). What is ESG? Deloitte. Retrieved December 5, 2022, from https://www2.deloitte.com/hu/en/pages/energy-and-resources/articles/e.html

Diether, K. B., Malloy, C. J., & Scherbina, A. (2002). Differences of Opinion and the Cross Section of Stock Returns. *The Journal of Finance*, *57*(5), 2113–2141.

Dillenburg, S., Greene, T., & Erekson, H. (2003). Approaching Socially Responsible Investment with a Comprehensive Ratings Scheme: Total Social Impact. *Journal of Business Ethics*, *43*(3), 167–177.

Doyle, T. M. (2018). Ratings That Don't Rate. Washington D.C. Retrieved November 4, 2022, from https://accfcorpgov.org/wp-content/uploads/2018/07/ACCF_RatingsESGReport.pdf

Dreman, D. N., & Berry, M. A. (1995). Analyst Forecasting Errors and Their Implications for Security Analysis. *Financial Analysts Journal*, *51*(3), 30–41.

Drempetic, S., Klein, C., & Zwergel, B. (2020). The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review. *Journal of Business Ethics*, *167*(2), 333–360.

Dyck, A., Morse, A., & Zingales, L. (2006). Who Blows the Whistle on Corporate Fraud? *The Journal of Finance*, *65*(6), 2213–2253.

Escrig-Olmedo, E., Fernández-Izquierdo, M. A., Ferrero-Ferrero, I., Rivera-Lirio, J. M., & Muñoz-Torres, M. J. (2019). Rating the Raters: Evaluating how ESG Rating Agencies Integrate Sustainability Principles. *Sustainability*, *11*(3), 1–16.

European Parliament and of the Council. (2014). Directive 2014/95/EU of the European Parliament and of the Council of 22 October 2014. *L 330*, 1–9. Official Journal of the European Union.

Fama, F. E., & French, K. R. (1992). The Cross-Section of Expected Stock Returns. *The Journal of Finance*, *47*(2), 427–465.

Fish, A. J. M., Kim, D. H., & Venkatraman, S. (2019). The ESG Sacrifice. Retrieved December 5, 2022, from http://dx.doi.org/10.2139/ssrn.3488475

Fisher, K. (2019). November 1867: The Invention of the Stock Ticker. Ken Fisher. Retrieved November 4, 2022, from https://www.kenfisher.com/market-history/november-1867-the-invention-of-the-stock-ticker

Frankel, R., & Li, X. (2004). Characteristics of a firm's information environment and the information asymmetry between insiders and outsiders. *Journal of Accounting and Economics*, *37*(2), 229–259.

Gell, S. (2011). *Determinants of Earnings Forecast Error, Earnings Forecast Revision and Earnings Forecast Accuracy* [Doctoral dissertation, University of Cologne].

Giese, G., Lee, L.-E., Melas, D., Nagy, Z., & Nishikawa, L. (2019). Foundations of ESG Investing: How ESG Affects Equity Valuation, Risk, and Performance. *The Journal of Portfolio Management*, *45*(5), 69–83.

Gilbert, M. (2021). The Rising Cost of Investing Responsibly. Bloomberg Opinion. Retrieved November 4, 2022, from https://www.bloomberg.com/opinion/articles/2019-01-11/the-rising-cost-of-esg-and-socially-responsible-investing?leadSource=uverify%20wall

Givoly, D., & Lakonishok, J. (1980). Financial Analysts' Forecasts of Earnings. *Journal of Banking and Finance*, *4*(2), 221–233.

Global Sustainable Investment Alliance (GSIA). (2021). Global Sustainable Investment Review 2020. Retrieved November 4, 2022, from http://www.gsi-alliance.org/wp-content/uploads/2021/08/GSIR-20201.pdf

Google Trends. (2022). ESG. Google Trends. Retrieved December 10, 2022, from https://trends.google.com/trends/explore?date=today%205-y&q=esg

Governance & Accountability Institute. (2022). Sustainability Reporting in Focus. Governance & Accountability Institute. Retrieved December 12, 2022, from http://www.ga-institute.com/fileadmin/ga_institute/images/FlashReports/2022/G_A-2022-Sustainability_Trends_Report.pdf?vgo_ee=fmIveivJ13fn9Ym%2FtP4nyNSYFmrMikCwlKFARSZoYAo%3D

Graham, B., & Dodd, D. L. (1934). *Security Analysis* (1st).

Graham, B., & Dodd, D. L. (2009). *Security Analysis* (6th).

Graham, J. R. (1999). Herding among investment newsletters: Theory and evidence. *The Journal of Finance*, *54*(1), 237–268.

Graham, J. R., Harvey, C. R., & Rajgopal, S. (2005). The Economic Implications of Corporate Financial Reporting. *Journal of Accounting and Economics*, *40*(1-3), 3–73.

Groysberg, B., Healy, P., & Chapman, C. (2008). Buy-Side vs. Sell-Side Analysts' Earnings Forecasts. *Financial Analysts Journal*, *64*(4), 25–39.

Han, B. H., & Manry, D. (2000). The Implications of Dispersion in Analysts' Earnings Forecasts for Future ROE and Future Returns. *Journal of Business Finance & Accounting*, *27*(1-2), 99–125.

Hankamper-Vandebulcke, N. (2021). Non-financial Reporting Directive. Retrieved February 3, 2023, from https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/654213/EPRS_BRI(2021)654213_EN.pdf

Healy, P. M., & Palepu, K. (2001). Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. *Journal of Accounting and Economics*, *31*(1-3), 405–440.

Henisz, W., Koller, T., & Nuttall, R. (2019). Five ways that ESG creates value. Retrieved November 5, 2022, from http://dln.jaipuria.ac.in:8080/jspui/bitstream/123456789/2319/1/Five-ways-that-ESG-creates-value.pdf

Hespenheide, E. J., & Koehler, D. A. (2013). Disclosure of long-term business value – What matters? Retrieved December 15, 2022, from https://www2.deloitte.com/content/dam/insights/us/articles/disclosure-of-long-term-business-value/DUP150_Reporting_What_Matters.pdf

Higgins, H. N., & Saito, Y. (2007). Do Analysts Play an Information Intermediary Role? Evidence from High and Low Intangible Firms. *SSRN*, 1–28.

Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to distraction: Extraneous events and underreaction to earnings news. *The Journal of Finance*, *64*(5), 2289–2325.

Hong, H., & Kubik, J. D. (2003). Analyzing the Analysts: Career Concerns and Biased Earnings Forecasts. *Journal of Finance*, *58*(1), 313–351.

Hong, H., Kubik, J. D., & Solomon, A. (2000). Security Analysts' Career Concerns and Herding of Earnings Forecasts. *The RAND Journal of Economics*, *31*(1), 121–144.

Hope, O.-K. (2003). Accounting Policy Disclosures and Analysts' Forecasts. *Contemporary Accounting Research*, *20*(2), 295–321.

Hou, J., Zhao, S., & Yang, H. (2020). Individual analysts, stock return synchronicity and information efficiency. *International Review of Financial Analysis*, *71*, 1–39.

Hwang, L. S., Jan, C.-L., & Basu, S. (2014). Loss Firms and Analysts's Earnings Forecast Errors. *The Journal of Financial Statement Analysis*, *1*(2), 18–30.

Imhoff, E. A., & Lobo, G. J. (1992). The Effect of Ex Ante Earnings Uncertainty on Earnings Response Coefficients. *The Accounting Review*, *67*(2), 427–439.

ISS ESG. (2022). ESG Corporate Rating. Retrieved January 22, 2023, from bit.ly/3QYAfiL

Jacobson, T. C. (1997). *From Practice to Profession: A History of the Financial Analysts Federation and the Investment Profession*.

Jennings, R. H. (1985). *Reaction of Financial Analysts to Management Earnings Forecasts*.

Johnson, T. C. (2004). Forecast Dispersion and the Cross Section of Expected Returns. *The Journal of Finance*, *59*(5), 1957–1978.

Jones, C. P., & Litzenberger, R. H. (1970). Quarterly Earnings Reports and Intermediate Stock Price Trends. *The Journal of Finance*, *25*(1), 143–148.

Kempf, A., & Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, *13*(5), 908–922.

Kendall, M. G., & Hill, A. B. (1953). The Analysis of Economic Time-Series-Part I: Prices. *Journal of the Royal Statistical Society*, *116*(1), 11–34.

Ketchum, M. D. (1967). Is financial analysis a profession? *Financial Analysts Journal*, *23*(6), 33–37.

Keynes, J. M. (2018). *The General Theory of Employment, Interest, and Money*.

Kimbrough, M., Wang, X., Wei, S., & Zhang, J. (2022). Does voluntary ESG reporting resolve disagreement among ESG rating agencies? [Forthcoming]. *European Accounting Review*.

Klettke, T. (2013). *New Determinants of Analysts' Earnings Forecast Accuracy* [Doctoral dissertation, University of Cologne].

Knorr-Cetina, K. (2011). Financial Analysis. In C. Camic, N. Gross, & M. Lamont (Eds.), *Social Knowledge in the Making* (pp. 405–441).

Koller, T., Goedhart, M., & Wessels, D. (2020). *Valuation: Measuring and Managing the Value of Companies* (7th).

Kotsantonis, S., & Serafeim, G. (2019). Four Things No One Will Tell You About ESG Data. *Journal of Applied Corporate Finance*, *31*(2), 50–58.

Kross, W., Ro, B., & Schroeder, D. (1990). Earnings Expectations: The Analysts' Information Advantage. *The Accounting Review*, *65*(2), 461–476.

Krueger, P., Sautner, Z., Tang, D. Y., & Zhong, R. (2021). The Effects of Mandatory ESG Disclosure Around the World. (754), 1–66.

LaBella, M. J., Sullivan, L., Russel, J., & Novikov, D. (2019). The Devil is in the Details: The Divergence in ESG Data and Implications for Responsible Investing. Retrieved January 2, 2023, from https://bit.ly/3WY6n7N

Lang, M. H., & Lundholm, R. J. (1996). Corporate Disclosure Policy and Analyst Behavior. *The Accounting Review*, *71*(4), 467–492.

Larcker, D. F., Pomorski, L., Tayan, B., & Watts, E. M. (2022). ESG Ratings: A Compass without Direction [Forthcoming].

Leins, S. (2018). *Stories of Capitalism: Inside the Role of Financial Analysts*.

Lu, S. (2021). Algorithmic Opacity, Private Accountability, and Corporate Social Disclosure in the Age of Artificial Intelligence. *Vanderbilt Journal of Entertainment & Technology Law*, *23*(1), 99–159.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417.

Mars, F. (1998). *Wir sind alle Seher: Die Praxis der Aktienanalyse*.

McClure, B. (2022). Earnings Forecasts: A Primer. Investopedia. Retrieved December 18, 2022, from https://www.investopedia.com/articles/stocks/06/earningsforecasts.asp

McDonald, C. L. (1973). An Empirical Examination of the Reliability of Published Predictions of Future Earnings. *The Accounting Review*, *48*(3), 502–510.

McKibben, W. (1972). Econometric Forecasting of Common Stock Investment Returns: A New Methodology using Fundamental Operating Data. *The Journal of Finance*, *27*(2), 371–380.

Morningstar. (2021). Enterprise Sustainability Report. Retrieved November 25, 2022, from https://assets.contentstack.io/v3/assets/blt4eb669caa7dc65b2/blt3ad3261518bad875/2021_Enterprise_Sustainability_Report.pdf

MSCI. (2019). MSCI to Strengthen Climate Risk Capability with Acquisition of Carbon Delta. Retrieved November 26, 2022, from https://ir.msci.com/node/15706/pdf

MSCI. (2022a). ESG Ratings Key Issue Framework. Retrieved November 26, 2022, from https://www.msci.com/ousolutions/esg-investing/esg-ratings/esg-ratings-key-issue-framework

MSCI. (2022b). MSCI ESG Ratings Methodology. Retrieved November 4, 2022, from https://www.msci.com/documents/1296102/21901542/ESG-Ratings-Methodology-Exec-Summary.pdf

MSCI ESG Research. (2022). SEC Investment Adviser Firm Summary - Part 2 Brochures. Retrieved November 26, 2022, from https://files.adviserinfo.sec.gov/IAPD/Content/Common/crd_iapd_Brochure.aspx?BRCHR_VRSN_ID=762070

Nagy, Z., Kassam, A., & Lee, L.-E. (2016). Can ESG Add Alpha? An Analysis of ESG Tilt and Momentum Strategies. *The Journal of Investing*, *25*(2), 113–124.

Nagy, Z., Lee, L.-E., & Giese, G. (2020). ESG Ratings: How the Weighting Scheme Affected Performance. MSCI Blog. Retrieved December 18, 2022, from https://www.msci.com/www/blog-posts/esg-ratings-how-the-weighting/01944696204

Ninety One. (2022). Ninety One survey finds European fund industry overly reliant on ESG scores. Ninety One. Retrieved December 7, 2022, from https://ninetyone.com/en/united-states/newsroom/ninety-one-survey-finds-european-fund-industry-overly-reliant-on-esg-scores

Nomura. (2022). How Artificial Intelligence is Reshaping ESG Ratings. Retrieved December 16, 2022, from https://d1qfwzw6aggd4h.cloudfront.net/about/ESG-Evaluation-using-AI-A4.pdf

Novethic. (2014). Overview of ESG Rating Agencies. Retrieved November 25, 2022, from https://www.novethic.fr/fileadmin/user_upload/tx_ausynovethicetudes/pdf_complets/2014_Overview-of-ESG-rating-agencies.pdf

Poon, M. (2012). Rating Agencies. In K. K. Cetina & A. Preda (Eds.), *The Oxford Handbook of The Sociology of Finance* (pp. 407–459). Oxford University Press.

Prall, K. (2021). ESG Ratings: Navigating Through the Haze. CFA Institute Blogs. Retrieved January 4, 2023, from https://blogs.cfainstitute.org/investor/2021/08/10/esg-ratings-navigating-through-the-haze/

Preda, A. (2006). Socio-Technical Agency in Financial Markets: The Case of the Stock Ticker. *Social Studies of Science*, *36*(5), 753–782.

Principles for Responsible Investment (PRI). (2022). Annual Report. Retrieved December 5, 2022, from https://dwtyzx6upklss.cloudfront.net/Uploads/b/f/m/pri_annual_report_2022_689047.pdf

Raghunandan, A., & Rajgopal, S. (2022). Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies*, *27*, 822–863.

Refinitiv. (2022a). About Us. Refinitiv. Retrieved November 28, 2022, from https://www.refinitiv.com/en/about-us#what-we-do

Refinitiv. (2022b). Environmental, Social and Governance Scores from Refinitiv. Retrieved November 28, 2022, from https://www.refinitiv.com/content/dam/marketing/en_us/documents/methodology/refinitiv-esg-scores-methodology.pdf

Refinitiv. (2022c). Refinitiv ESG company scores. Refinitiv. Retrieved November 28, 2022, from https://www.refinitiv.com/en/sustainable-finance/esg-scores

Reuters. (2018). Thomson Reuters closes deal with Blackstone. Reuters. Retrieved November 28, 2022, from https://www.reuters.com/

article/us-thomsonreuters-m-a-blackstone/thomson-reuters-closes-deal-with-blackstone-idUSKCN1MB3PY
Riles, A. (2006). *Documents: Artifacts of modern knowledge*. University of Michigan Press.
Riles, A. (2011). *Collateral knowledge: Legal reasoning in the global financial markets*. University of Chicago Press.
Ryan, P., & Taffler, R. J. (2004). Are Economically Significant Stock Returns and Trading Volumes Driven by Firm-specific News Releases? *Journal of Business Finance & Accounting*, *31*(1& 2), 49–82.
Scalet, S., & Kelly, T. F. (2010). CSR Rating Agencies: What is Their Global Impact? *Journal of Business Ethics*, *94*, 69–88.
Scharfstein, D. S., & Stein, J. C. (1990). Herd behavior and investment. *The American economic review*, 465–479.
Securities and Exchange Commission (SEC). (2010). Analyzing Analyst Recommendations. SEC. Retrieved November 5, 2022, from https://www.sec.gov/tm/reportspubs/investor-publications/investorpubsanalystshtm.html
Securities and Exchange Commission (SEC). (2022). SEC Proposes to Enhance Disclosures by Certain Investment Advisers and Investment Companies About ESG Investment Practices. Retrieved December 12, 2022, from https://www.sec.gov/news/press-release/2022-92
Sheng, X., & Thevenot, M. (2012). A new measure of earnings forecast uncertainty. *Journal of Accounting and Economics*, *53*(1-2), 21–33.
Sindreu, J., & Kent, S. (2018). Why It's So Hard to Be an 'Ethical' Investor. The Wall Street Journal. Retrieved November 4, 2022, from https://www.wsj.com/articles/why-its-so-hard-to-be-an-ethical-investor-1535799601
Sipiczki, A. (2022). A critical look at the ESG market. Retrieved December 16, 2022, from https://www.ceps.eu/wp-content/uploads/2022/04/PI2022-15_A-critical-look-at-the-ESG-market.pdf
S&P Global. (2020). How can AI help ESG investing? S&P Global. Retrieved December 16, 2022, from https://www.spglobal.com/en/research-insights/articles/how-can-ai-help-esg-investing
S&P Global. (2022). S&P Global ESG Scores – Methodology. Retrieved January 22, 2023, from https://www.spglobal.com/esg/documents/sp-global-esg-scores-methodology-2022.pdf
Stack Exchange. (2023). Exclusion of Utilities and Financials in Magic Formula. Quantitative Finance Stack Exchange. Retrieved January 30, 2023, from https://quant.stackexchange.com/questions/41192/exclusion-of-utilites-and-financials-in-magic-formula
State Street Global Advisors. (2019). The ESG Data Challenge. Retrieved January 4, 2023, from https://www.ssga.com/investment-topics/environmental-social-governance/2019/03/esg-data-challenge.pdf
Stickel, S. E. (1990). Predicting individual analyst earnings forecasts. *Journal of Accounting Research*, *28*(2), 409–417.
Sustainalytics. (2020). The ESG Risk Rating: Frequently Asked Questions – For Companies. Retrieved November 25, 2022, from shorturl.at/kopX3
Sustainalytics. (2021). ESG Risk Ratings – Methodology Abstract. Retrieved November 25, 2022, from https://connect.sustainalytics.com/esg-risk-ratings-methodology
Sustainalytics. (2022). About Us. Sustainalytics. Retrieved November 25, 2022, from https://www.sustainalytics.com/about-us
Swaminathan, S. (1991). The Impact of SEC Mandated Segment Data on Price Variability and Divergence of Beliefs. *The Accounting Review*, *66*(1), 23–41.
Tang, D. Y., Yan, J., & Yao, C. Y. (2022). The Determinants of ESG Ratings: Rater Ownership Matters. *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI – ESSEC*.
The Economist. (2019). The London Stock Exchange buys Refinitiv. The Economist. Retrieved November 28, 2022, from https://www.economist.com/finance-and-economics/2019/08/01/the-london-stock-exchange-buys-refinitiv
The SustainAbility Institute. (2020). Rate the Raters 2020: Investor Survey and Interview Results. Retrieved December 7, 2022, from https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/sustainability-ratetheraters2020-report.pdf
The SustainAbility Institute. (2022). Costs and Benefits of Climate-Related Disclosure Activities by Corporate Issuers and Institutional Investors. Retrieved January 2, 2023, from https://www.sustainability.com/globalassets/sustainability.com/thinking/pdfs/2022/costs-and-benefits-of-climate-related-disclosure-activities-by-corporate-issuers-and-institutional-investors-17-may-22.pdf
Tomoko, F., & Kyoko, N. (2022). Japan Introduces Mandatory ESG Disclosures For Public Companies. Morgan Lewis. Retrieved February 3, 2023, from https://www.morganlewis.com/pubs/2022/12/japan-introduces-mandatory-esg-disclosures-for-public-companies
UN PRI and ICGN. (2018). Investor Agenda For Corporate Esg Reporting. Retrieved December 12, 2022, from https://www.unpri.org/download?ac=6181
Wansleben, L. (2012). Financial Analysts. In K. K. Cetina & A. Preda (Eds.), *The Oxford Handbook of The Sociology of Finance* (pp. 407–459). Oxford.
Welch, I. (2000). Herding among security analysts. *Journal of Financial Economics*, *58*(3), 369–396.
Wild, D. (2017). ESG Integration Approach. In C. Institute, C. S. Switzerland, & S. S. Finance (Eds.), *Handbook on Sustainable Investments* (pp. 54–59).
Winroth, K., Blomberg, J., & Kjellberg, H. (2010). Enacting Overlapping Markets. *Journal of Cultural Economy*, *3*(1), 3–18.
Yu, F. (2008). Analyst coverage and earnings management. *Journal of Financial Economics*, *88*(2), 245–271.
Zhang, H. (2021). It's Time for Private Companies to Come Clean on ESG. Institutional Investor. Retrieved January 2, 2023, from https://www.institutionalinvestor.com/article/b1v8zb72tjzcy3/It-s-Time-for-Private-Companies-to-Come-Clean-on-ESG
Zmijewski, M. E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, *22*, 59–82.
Zwiebel, J. (1995). Corporate Conservatism and Relative Compensation. *Journal of Political Economy*, *103*(1), 1–25.

Junior Management Science 9(3) (2024) 1805-1835

Junior Management Science

www.jums.academy
ISSN: 2942-1861

Development of a Cost Optimal Predictive Maintenance Strategy

Christoph Weeber

Technical University of Munich

Abstract

Maintenance costs account for a significant share of operating expenses. Selecting the optimal maintenance strategy for each application is crucial to optimize operational processes and minimize MRO spending. In recent years, Machine Learning has become popular for analyzing large amounts of data and improving decision-making in various industries. This yields great potential in the field of Predictive Maintenance. In this thesis, a methodology to determine and compare the average maintenance costs per cycle for Reactive, Preventive, and Predictive Maintenance, as well as a Reference Case is developed. This cost comparison methodology is then applied to a realistic example of a fleet of ten aircraft. Unlike previous research, this thesis combines all aspects in one approach, from Machine Learning algorithm selection and RUL prediction, to the maintenance cost comparison based on a fleet of aircraft. The NASA CMAPSS jet engine dataset is used as an example. Results suggest that maintenance costs per cycle for Predictive Maintenance are 36.0 % lower than for Preventive Maintenance and 88.3 % lower compared to Reactive Maintenance. In general, this thesis serves as a guideline that highlights the necessary steps to determine the cost-optimal maintenance strategy for an application.

Keywords: machine learning algorithm; NASA CMAPSS dataset; optimal maintenance strategy; predictive maintenance; preventive maintenance; reactive maintenance

1. Introduction

Production systems and machines are becoming increasingly more complex. Failures can be extremely costly, not only in terms of expensive spare parts and for conducting the maintenance tasks themselves, but also due to lost profit caused by unavailable production systems and machines. In today's globalized world, increasing competition is forcing companies to optimize wherever possible to stay competitive. Thus, long machine downtime can be particularly detrimental.

According to the IATA Airline Maintenance Cost Executive Commentary, global aircraft Maintenance, Repair and Overhaul (MRO) spending in 2019 was $91 billion, which is expected to increase to approx. $118 billion in 2030 as the worldwide passenger and cargo volume are predicted to grow each year. This represents 11.2% of the airline's total operating costs. Broken down to a single aircraft, this accumulates to approx. $3.3 million MRO spending per aircraft per year. Nearly half of those costs (approximately 43%) can be attributed to engine maintenance (IATA, 2021). A large portion of the ticket price for airline customers, thus is made up of aircraft and engine maintenance costs.

This percentage can be even higher in the manufacturing industry where maintenance costs can account for up to 40% of total operating expenses (Haroun, 2015). Maintenance costs have also been increasing progressively in the past decades due to more complex machines, which makes their impact on business performance even more significant (Mobley, 2002). According to maintenance management effectiveness surveys, around 33% of the maintenance costs are wasted due to unnecessary over maintenance or improper maintenance. This equates to an annual loss of approximately $60 billion in the U.S. alone, considering the total spending on maintenance of approximately $200 billion in the U.S. each year (Mobley, 2002). This clearly shows that maintenance is not only a significant cost driver, but also emphasizes the necessity to optimize the maintenance strategy to realize the drastic potential savings.

DOI: https://doi.org/10.5282/jums/v9i3pp1805-1835

Due to the stringent safety requirements in aviation, the maintenance of aircraft is strictly regulated. It must be noted that safety is and should always be the primary concern in this industry.

There are three different maintenance strategies with distinct advantages and disadvantages which can be classified into the following categories (Lee & Scott, 2009):

Reactive Maintenance, or Corrective Maintenance, encompasses all strategies where the system or machine is run until it breaks down. As the name suggests, all maintenance is reactive to problems once they occur. Depending on the application, this is usually the costliest approach due to high inventory holding costs to have all necessary spare parts available, as well as high personnel costs for the technicians to conduct the repair on short notice. However, this approach might be most suitable for components that can't be maintained and are easy to fix or replace, such as a lightbulb (Mobley, 2002; Sirvio, 2015).

A more sophisticated approach is called **Preventive Maintenance**, which includes all timedriven strategies by considering the Mean-Time-To-Failure (MTTF). The required maintenance task is scheduled early enough to prevent component or system failure. However, every machine is different and the MTTF can vary greatly depending on various circumstances, such as the specific operating conditions. By simply considering the same MTTF for all devices of the same type, most devices are over-maintained, which results in higher maintenance costs than necessary (Mobley, 2002).

The third approach, **Predictive Maintenance**, is condition driven and considers the actual state of the machine when determining the maintenance time. Based on past and current machine data, a machine learning algorithm can be used to determine a variety of system parameters such as: detection of anomalies; isolation/diagnosis of occurring failures; prediction of the health state of the system; and estimation of its Remaining Useful Life (RUL). The RUL estimate can be used to schedule the required maintenance tasks optimally, e.g., when the production is running at a lower rate (Mobley, 2002; Sirvio, 2015).

Depending on the application, some approaches may be more advantageous than others, which is why an individual optimal maintenance strategy for each specific application has to be considered. Especially due to the advancements in the field of machine learning and increasing computing power, Predictive Maintenance is becoming more and more relevant in various industries.

This thesis aims to provide insights into the steps required to determine the cost-optimal Maintenance strategy. The NASA Commercial Modular Aero-Propulsion Simulation (CMAPSS) Turbofan Engine Data Set is used to conduct the research and shall serve as an example (NASA, 2023).

In Chapter 2, a primary literature review shall lay out the findings of available research and pinpoint where information is still missing. Chapter 3 then provides background information into the different Maintenance Strategies, the specifics of Aircraft and Jet Engine Maintenance, as well as the concept of Machine Learning. The CMAPSS dataset is analyzed and prepared in Chapter 4. In Chapter 5, a Machine Learning model is proposed to determine the average maintenance costs per cycle of different maintenance strategies. The application of the developed method as well as the results are provided in Chapter 6. In a further step (Chapter 7), the method is applied to a realistic example of a fleet of ten aircraft. The Machine learning model described in Chapter 5, and the methodology proposed in Chapters 6 and 7 are implemented in Python, using the scikit-learn Machine Learning library. In short, this thesis shall take all necessary steps from Machine Learning algorithm selection to RUL prediction, up to the cost comparison of different maintenance strategies based on a fleet of aircraft into account.

2. Literature Review

This literature review shall describe the focus of previous research to highlight where information is still missing to show how this thesis will contribute to filling the gaps. It is structured in the following way:

- First, different sources which compare the performance of Machine Learning algorithms suitable for prognostics are presented.
- In the next step, different papers which focus on integrating such algorithms into prognostics and RUL prediction are discussed.
- Research is presented where RUL prediction is used during maintenance planning.
- In the final step, literature is discussed where those concepts are applied to a realistic maintenance framework of an aircraft fleet.

It must be noted that this literature review only contains papers and other forms of literature which have been decided to be of high relevant for this thesis. There are also numerous other sources available that are not mentioned here.

Machine Learning and especially Predictive Maintenance is a relatively new field. Due to increasing computing power and ever more powerful machine learning algorithms, it has become of interest for various applications across many industries. For the past ten years, ongoing research has been conducted to expand those possible applications further, such as Cline et al. (2017) and Tiddens et al. (2020).

An essential prerequisite for Predictive Maintenance is prognostics, the overarching principle of RUL prediction. When Machine Learning was first used in prognostics, many models were based only on a single parameter, or feature, to determine the system degradation and the RUL, see Junqiang et al. (2014). Considering an aircraft engine as an example, the Exhaust Gas Temperature Margin (EGTM) was a standard parameter used in those models. However, due to the high complexity of such systems, multiple-parameter models were proposed to improve the prediction. In Junqiang et al.

(2014), a framework to fuse the information of multiple parameters was developed, using a Kalman filter to achieve the RUL prediction. Using multiple sensors as input is a common practice by now, so most of the following literature already incorporates information fusion.

To build a suitable Machine Learning prognostics model and predict the RUL, it must be differentiated between physics - based and data-driven approaches. Further details about physics-based and data-driven approaches and their differences will be described in Chapter 5. Deciding which approach and algorithm is most suitable for the given problem can be complex. To solve this issue, different papers provide a sound basis by evaluating the differences between various data-driven and physics-based Machine Learning algorithms, as well as their pros and cons (An et al., 2015; Carvalho et al., 2019; Silvestrin et al., 2019; Singh et al., 2020).

For example, in An et al. (2015), a fatigue crack growth example is used to evaluate the performance of different algorithms, such as Neural Networks and Gaussian Process Regression for the data-driven approaches. For the physics-based approaches, Bayesian Methods and Particle Filters are considered. For each described algorithm, advantages and disadvantages are provided. The results suggest that Neural Networks are advantageous, particularly if the model is very complex or if high noise levels within the data may obstruct the model. On the other hand, Silvestrin et al. (2019) compares Neural Networks with more traditional algorithms, such as Decision Trees, Random Forests, and KNNs, and concludes that traditional algorithms may be more advantageous if only very limited training data is available.

Multiple papers propose even more complex algorithms for RUL predictions with a higher performance than traditional algorithms and Neural Networks. For example, H. Li et al. (2020) and X. Li et al. (2018) and dePater et al. (2022) propose a Convolutional Neural Network for RUL prediction, using the CMAPSS Turbofan dataset as the underlying training and testing data.

To set the results of different algorithms across multiple studies into perspective, Vollert and Theissler (2021) compare the performance of the proposed algorithms in 81 different publications, which all use the CMAPSS dataset as a basis. Performance is measured in terms of Root Mean Squared Error (RMSE); see Chapter 5. When considering the subdataset FD001, Vollert and Theissler (2021) point out that different types of Artificial Neural Networks show a high performance (low RMSE values). A Convolutional Neural Network, or CNN, sets the lower baseline (RMSE = 8), followed by a Long-Short Term Memory, or LSTM (RMSE = 11). Tree-based algorithms and the Multi-Layer Perceptron also show promising results. It is essential to consider the lowest RMSE of each algorithm and the spread of the same algorithm across multiple papers. As it turns out, the RMSE ranges from around 8 to 18 for the CNN and around 11 to 23 for the LSTM. Conversely, the spread is much lower for the Multi-Layer Perceptron (MLP), ranging from around 13 to 16.

Still, the choice of the optimal algorithm strongly depends on the specific application. Some types of algorithms may outperform other algorithms in some applications, but show less promising results in others. As described in Chapter 5, this thesis includes a preliminary algorithm comparison to determine the optimal algorithm for the provided dataset. The described literature serves as a guideline for selecting the most promising algorithms for the provided dataset, as well as the ranges for the hyperparameter tuning. Due to the high fluctuations of performance scores across different studies, such a preliminary comparison is always recommended.

In order to incorporate a Machine Learning algorithm into RUL prediction and ultimately into Predictive Maintenance, several health monitoring techniques for different components have been proposed. Many studies provide detailed insights into strategies for integrating sensors into aircraft structures to continuously monitor degradation and faults. For example, Diamanti and Soutis (2010) describe techniques for continuously monitoring composite aircraft structures to increase operational safety. Ignatovich et al. (2013) shows that it is possible to continuously estimate fatigue damage of metal structures in aircraft and incorporate this into aircraft structural health monitoring. Zhao et al. (2007) and Ihn and Chang (2014) both incorporate piezoelectric sensors into different aircraft structures, such as the wing, to monitor hidden fatigue cracks and their growth size.

Most papers can be divided into two distinct categories: the first group mainly focuses on RUL prediction and which Machine Learning algorithms are particularly suitable for the given problem. The other group assumes that the prognostics information or the model to determine the system degradation is already known. Therefore, the second group mainly focuses on maintenance optimization or operations planning. Only a few papers combine those two categories and provide a complete framework incorporating both aspects (Gilabert et al., 2017; Nguyen & Medjaher, 2019).

Wang et al. (2017) and dePater et al. (2022) even take this one step further. Wang et al. (2017) proposes a physics-based prognostics framework to determine the crack size evolution and then uses this information to apply the framework to a realistic simulation of maintenance processes of an entire fleet of aircraft. Based on this realistic scenario, the resulting costs of Predictive Maintenance are then compared with two other maintenance strategies, Scheduled Maintenance, and Threshold Based Maintenance. Results show that maintenance costs can be reduced by employing Predictive Maintenance compared with both other strategies. However, unlike most other papers, the “future system reliability” is introduced as a prognostic index, while most papers use RUL. Also, since the paper builds on a physics-based model, it does not rely on acquired data but assumes that the crack size evolution can be described by a stochastic process. dePater et al. (2022) follows a similar approach, but a data-driven Convolutional Neural Network is used to predict the RUL of jet engines, relying on the CMAPSS dataset as the basis. The RUL prognostics framework is also integrated into maintenance planning, where an entire fleet of aircraft is simulated,

and the engine maintenance tasks and spare part orders are planned. However, the resulting maintenance costs are only compared for perfect and imperfect RUL prediction; other maintenance strategies, such as Preventive or Reactive Maintenance, are not taken into account.

This thesis aims to combine and extend existing research: a preliminary Machine Learning algorithm comparison and hyperparameter tuning shall help determine the optimal algorithm for the CMAPSS dataset. Based on those preliminary findings, a data-driven model is developed for RUL prediction. Using this model, a methodology to compare the maintenance costs of different maintenance strategies is developed. In the final step, a fleet of ten aircraft is considered to apply the methodology to a more realistic scenario. A similar approach was developed by dePater et al. (2022), however, only perfect and imperfect RUL prediction was considered, while Reactive and Preventive Maintenance were neglected.

3. Background Information

This thesis touches upon a variety of different topics. A brief introduction into maintenance strategies, aircraft maintenance, and Machine Learning shall provide the required background information.

3.1. Machine Learning

In most industries, there is a clear shift towards data-driven operations and a high reliance on big data to optimize production processes, maximize throughput and minimize occurring costs as much as possible. Machine Learning has become popular in dealing with large amounts of generated data and improving relevant decisions. New and ever more powerful algorithms and increasing computing power are essential prerequisites for highly accurate predictions by the algorithm. Especially in areas where conventional algorithms are not able to perform specific tasks, such as speech recognition, computer vision, or big data analysis, machine learning has a significant advantage (Hu et al., 2022; Silvestrin et al., 2019). This introduction shall provide relevant background information.

Machine learning algorithms use training data and extract relevant features to build a model without explicitly being programmed in a certain way. This resulting model can then use new, unseen input data to predict the corresponding output and make decisions without human interaction. Conventional algorithms, on the other hand, use the input data and then determine the output based on specified calculations. Especially if the data is very complex and multivariate, programming a conventional algorithm can be extremely difficult.

Machine Learning algorithms can perform a variety of different tasks, two of them being data classification and regression. Data classification is used to analyze and sort input data and classify each data point into a distinct category; for example, analyzing different pictures of animals and assigning them to categories such as “dog” or “cat” is considered as classification. Regression uses input data to predict the corresponding numerical output (Vollert & Theissler, 2021). One example is the prediction of house prices based on the number of bedrooms, living area, and house age. The different types of algorithms can be categorized into three main clusters (Lei et al., 2018):

- **Supervised Learning:** the training data contains inputs (each distinct input is also referred to as a feature) and the corresponding output (called a label). Supervised learning algorithms extract the relevant information from this data and learn which features are relevant to predict the corresponding label accurately.
- **Unsupervised Learning:** only the features are available; therefore, the algorithm must find a way to structure the features and find hidden patterns within the data itself.
- **Reinforcement Learning:** the algorithm tries to perform a specific task, e.g., driving an autonomous vehicle, and receives positive or negative feedback. The algorithm attempts to maximize these “rewards” and adapts its output accordingly.

A common phenomenon among many machine learning algorithms is called “overfitting”. During the training process, the model learns from the training data and adapts accordingly to predict the output from the input as accurately as possible. Suppose the algorithm picks up too much of the noise of the input data and learns from the noise instead of the more general underlying information. In that case, the model is not generalizing the input data well enough, which has to be prevented by countermeasures (Vollert & Theissler, 2021).

3.2. Maintenance Strategies

Maintenance, often referred to as “MRO” - Maintenance, Repair, and Overhaul - can be defined in a variety of ways, one being “any activity - such as tests, measurements, replacements, adjustments, and repairs - intended to retain or restore a functional unit in or to a specified state in which the unit can perform its required functions” (US Department of Defense, 2004). Maintenance has two major objectives: high availability of production equipment and low maintenance costs. This is an apparent contradiction because higher maintenance spending usually correlate with a lower chance of system failure (Deighton, 2016).

Three major maintenance strategies must be distinguished: Corrective Maintenance, Preventive Maintenance, and Predictive Maintenance (Lee & Scott, 2009). It must be noted that these categories may vary if other literature is considered.

3.2.1. Corrective Maintenance

The most straightforward strategy is Corrective Maintenance, also referred to as Run-ToFailure or Reactive Maintenance. This strategy aims to “identify, isolate, and rectify

a fault" (US Department of Defense, 2011) after it has occurred, which means the machine or component is repaired or replaced after failure; there are no maintenance tasks up to this point. This approach is rarely used in its proper form because, most of the time, at least some basic preventive tasks, such as machine lubrication or some forms of adjustment, are performed in specific time intervals (Sirvio, 2015).

Corrective Maintenance has advantages and disadvantages: it is straightforward to implement because no maintenance tasks must be planned, and no maintenance costs incur in advance. However, costs are typically significantly higher if an unplanned maintenance task occurs instead of a planned task: first, an unscheduled task may result in unplanned machine downtime and production loss, which decreases machine productivity. Also, a higher inventory of all major spare parts must be kept, thus increasing inventory holding costs. Outsourcing the spare part inventory may be possible by purchasing the parts whenever they fail; however, vendors typically charge additional premiums if the parts are delivered on short notice (Mobley, 2002; Sirvio, 2015).

The maintenance staff has to be kept on stand-by and ready whenever a failure occurs, thus increasing personal costs and even overtime costs if necessary (Mobley, 2002). Also, depending on the machine, a failure of certain components may result in additional damage, which also increases the cost for restoration.

Studies have shown that the overall maintenance costs of Reactive Maintenance can be up to three times higher than scheduling and performing the same tasks in advance before failure occurs (Mobley, 2002). Still, there are applications where this strategy is most suitable, such as easy-to-reach lightbulbs, batteries in some remote-control devices, or similar. Storing a sufficient supply of those spare parts incurs negligible inventory holding costs and labor costs for Maintenance (e.g., exchanging the lightbulb). Also, failure of such components is not considered a safety risk, unlike the failure of a fan blade of a jet engine.

3.2.2. Preventive Maintenance

On the other hand, Preventive Maintenance is a strategy where maintenance tasks are conducted based on elapsing system parameters, such as time, cycles, produced units, or driven miles. The conducted maintenance tasks vary greatly and depend on the machine or component: it can range from simple lubrication, oil change, or adjustments to very complex tasks such as disassembling an aircraft engine and inspecting each part individually (Lee & Scott, 2009).

The machine or component's MTTF is considered to determine the optimal maintenance interval. The MTTF can be determined for each machine or component by considering the failure rates. It usually follows a so-called bathtub curve: during the beginning, the number of failures is high because of defective parts that are not noticed before installation or if the installation is not conducted correctly. After this initial period, the curve stays relatively low for a more extended period until the first components start to fail due to wearing out (Mobley, 2002; Sirvio, 2015). Determining the optimal maintenance interval is described in more detail in Chapter 6.

One major problem of this approach is the assumption that all machines or components follow a uniform degradation behavior, i.e., fail at approximately the same time/cycle/mile. Each machine's environmental and operating condition can be considered to some extent, but only if the correlation between the operating condition and accelerated degradation is known. If a machine is subject to harsh weather, the time to the following maintenance task may be lowered by some factor. However, each machine is different due to different production variabilities, operating histories, and other factors that can hardly be considered when determining the optimal maintenance interval. This results in either over-maintained machines, which means maintenance tasks occur more often than necessary, or in machine failure (Mobley, 2002). Both cases result in higher maintenance costs, additional machine downtime, and reduced profitability, as shown in Figure 1: if the maintenance interval is low, the preventive costs (orange) are high due to unnecessarily conducted maintenance tasks. On the other hand, if the interval between maintenance tasks is too high, the repair costs due to machine failures are high (Lee & Scott, 2009; Mobley, 2002). The optimal interval is shown in green, with minimal total maintenance costs.

3.2.3. Predictive Maintenance

The third and most advanced approach is Predictive Maintenance, sometimes referred to as Condition-Based Maintenance. As opposed to Reactive or Preventive Maintenance, this strategy considers the actual machine condition to determine the optimal maintenance time for each machine individually (Mobley, 2002).

A prerequisite for Predictive Maintenance is prognostics: in a primary step, different machine parameters, such as temperature, pressure, vibration, or fuel flow, are monitored by a network of sensors. There are several non-destructive ways to record such parameters during operation (Diamanti & Soutis, 2010). The gathered sensor data is then analyzed to assess the current system condition and predict the time to failure, or RUL (Lei et al., 2018).

To evaluate the sensor data, Machine Learning algorithms are commonly used. The different failure modes that can occur and how each can be detected based on a change in the sensor values must be well understood when the model is developed. This is particularly the case for physics-based models, while data-driven approaches can extract some of this information themselves, see Data-driven approaches below. This knowledge is also essential when determining the types of sensors and mounting locations, as well as for the selection of the optimal Machine Learning algorithm (An et al., 2015; Hu et al., 2022).

Prognostic models can be assigned to three categories: data-driven, physics-based (or modelbased), and hybrid forms (Vachtsevanos et al., 2006).

Data-driven approaches usually require large amounts of previously recorded data to be trained effectively. Dur-

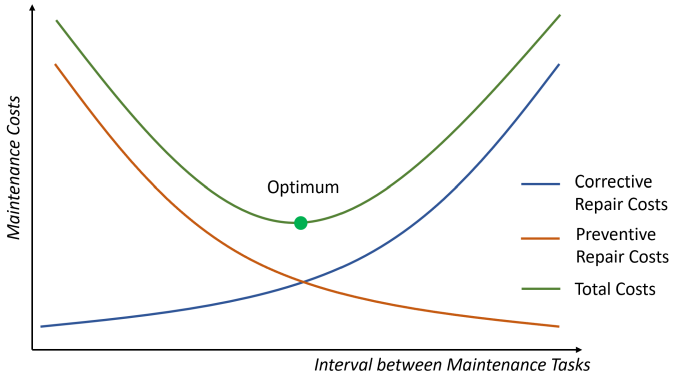


Figure 1: Representation of the optimal maintenance interval (Own graphic, derived from Dawotola et al. (2013))

ing this training, the Machine Learning algorithm learns and adapts to find patterns in the data. The resulting model can then be used to predict the failure mode, current system condition, or RUL during operation. Data-driven models are particularly suitable if the system is too complex to build a physical model or if understanding the correlation between sensor values and failure or RUL is too difficult.

The advantage of the data-driven model is that it extracts the relevant information from the available data itself without explicitly being programmed in a certain way. More information about Machine Learning and algorithm training is provided in Chapter 5.

Physics-based approaches can be used if no such data is available or if the degradation can be described with physical correlations. In this case, a physical model which explains the system as accurately as possible is developed. This physical model is then used to determine the current system state, future condition, and RUL from the fed sensor data. Usually, systems are very complex and must be approximated somehow, which may limit the performance of physics-based approaches. **Hybrid approaches** are a combination of both (Hu et al., 2022). In the following, only the data-driven approach is considered.

The integration of prognostics into maintenance planning is referred to as Predictive Maintenance. The predicted RUL can be used to determine when the maintenance task should be performed in the most cost-optimal way. This is usually as close to failure as possible, but before performance deteriorates too significantly or failure occurs. Also, machine utilization can be taken into account to minimize the disruption to the operation. Predictive Maintenance increases the safety and optimizes spare part usage by reducing the likelihood of unexpected failure. That way, the maintenance costs can be minimized while maximizing the machine's productivity. On the other hand, Predictive Maintenance can be difficult and costly to implement (Mobley, 2002; Tiddens et al., 2020).

3.3. Maintenance Cost Allocation

As stated above, maintenance costs have a significant effect on company finances. Studies suggest that maintenance costs can account for up to 40% of total company turnover, depending on the industry (Haroun, 2015).

As the complexity of machines in all industries has been increasing, maintenance has become even more costly. However, the exact amount is often difficult to assess. From a cost accounting perspective, tracing the maintenance costs directly to a cost driver is hard to implement. Usually, overheads are used for this cost allocation. If a company evaluates the potential of switching from a Preventive Maintenance strategy to a Predictive Maintenance strategy, estimating the cost/benefit can be particularly difficult. In order to evaluate this reliably, a sufficient amount of past data has to be available to determine the frequency and randomness of breakdowns, excessive fuel/energy consumption during operation, reduced throughput during operation, etc. (Haroun, 2015).

When considering aircraft engines as an example, the exact maintenance costs, also called Direct Maintenance Costs (DMC), and the engine's endurable Time On-Wing (TOW) vary greatly and depend on several criteria, such as the thrust rating of the engine, operational severity (e.g., if the aircraft is used for short- or long-range flights), the maturity of the engine and the operating conditions (such as ambient temperature or air quality) (Shannon & Ackert, 2011).

Generally, as described above, aircraft maintenance accounts for approximately $10 - 15\%$ of the airline's operating expenses, and of those, around 43% can be attributed to engine maintenance (IATA, 2021). Considering the average MRO spending per aircraft per year of \$3.3 million and assuming that 43% of those costs are attributed to engine maintenance, this accounts for \$1.42 million per aircraft for engine maintenance per year. Reducing engine maintenance costs by only 1% will result in cost savings of $\$14,200$ per aircraft per year. Although evaluating the optimal maintenance strategy and the cost benefit in applications such as aircraft engines requires effort and know-how, significant savings can be realized.

3.4. Introduction to Aircraft and Jet Engine Maintenance

Before explaining the underlying model of this thesis and comparing the incurred maintenance costs per cycle of individual maintenance strategies, a brief introduction to aircraft and jet engine maintenance shall provide the necessary background information. It is essential to understand how and when maintenance costs occur to minimize them by selecting the optimal maintenance strategy.

The FAA requires regular maintenance checks and repairs to ensure that the safety requirements of the aircraft are always met. Therefore, the aircraft manufacturer develops specific maintenance check schedules and corresponding intervals for each aircraft type. The exact interval between those checks can vary depending on several criteria, such as the aircraft type, accrued flight hours and cycles, and operating conditions. The individual checks can be divided into categories, also referred to as letter checks (Shannon & Ackert, 2010):

- **A-checks** are performed every 400-600 flight hours, or every 200-300 cycles, depending on the aircraft type. They are relatively easy to perform and require about $50-70$ person-hours. A typical aircraft undergoes an A-check at the hangar every 10 to 20 days (dePater et al., 2022). Common checklist items are inspecting interior and exterior surfaces with selected doors open and electrical checks (Department for Business Innovation and Skills, 2016; Shannon & Ackert, 2010).
- **B-checks** are performed every 6-8 months and are thus more thorough when the aircraft is grounded for around 2-3 days (Department for Business Innovation and Skills, 2016).
- **C-checks** are much more extensive than the described A - and B-checks and occur every 1220 months. During a C-check, most components and systems of the aircraft are inspected (Shannon & Ackert, 2010).
- **D-checks**, also referred to as "heavy maintenance checks", occur every 6-12 years and require the entire airplane to be dismantled. All individual parts are then thoroughly inspected and overhauled if necessary. The maintenance costs for a 747-400 lie between \$4.0 million and \$4.5 million (Department for Business Innovation and Skills, 2016; Shannon & Ackert, 2010).

Aircraft engines are the most complex part of an aircraft and require entirely different maintenance tasks than the airframe. The airframes' and the engines' maintenance intervals are usually synchronized as much as possible to reduce the time the aircraft is grounded.

There are three different objectives when the aircraft engine is maintained (Shannon & Ackert, 2011):

- **Operational:** to keep the engine in good operational condition.
- **Value Retention:** to reduce the engine's deterioration and maintain its value.
- **Regulatory Requirements:** set by regulatory authorities to ensure safety.

When the engine undergoes Maintenance, two elements must be considered, which are (Shannon & Ackert, 2011):

- **Performance Restoration:** during operation, many parts of the engine are exposed to high temperatures and extreme centrifugal forces, which deteriorate the engine's performance through erosion, fatigue, and residue accumulation. As the engine ages, the Exhaust Gas Temperature (EGT) increases, further accelerating the performance deterioration. If the EGT reaches critical levels, it is, therefore, necessary to dismantle the engine to inspect, repair, clean, or replace the necessary parts to restore performance.
- **Life Limited Parts Replacement:** many components throughout the engine, such as turbine blades, disks, or shafts, must be replaced after a certain number of cycles. A failure of those components could not be contained and may lead to a catastrophic incident. During engine maintenance, the parts which approach their life limit are replaced. Depending on how the engine is operated, some components may never have to be replaced during the entire engine life, mainly if only long-range flights are conducted.

For the past decades, jet engines were usually maintained based on a fixed schedule, often referred to as "hard time interval", without considering the current condition of the engine. Due to the availability of highly accurate sensors and Machine Learning algorithms, condition monitoring is becoming more relevant in engine maintenance, where it is usually referred to as "Engine Trend Monitoring". Machine Learning models analyze the information provided by the sensors to determine the degradation and RUL of the engine. Based on this information, the required maintenance task is scheduled in the most cost-optimal way. This can be particularly suitable for the first of the two described cases to detect performance degradation well in advance. Life-limited parts must still be replaced based on a fixed interval; thus, RUL prediction is not applicable here (Shannon & Ackert, 2011).

Jet engines are regarded as highly complex machines with many different failure modes, some of which are very difficult to predict in advance. The system, in this case, the jet engine, can be described as a macroscopic or a microscopic system. The macroscopic system considers individual modules (High-Pressure Compressor (HPC), Low-Pressure Compressor (LPC), High-Pressure Turbine (HPT), Low-Pressure Turbine (LPT)) as a whole. Degradation of each module, e.g., HPC degradation, can be detected by different sensors throughout the module. On the other hand, the microscopic system also takes all individual components into account, such as turbine or compressor blades, vanes, and

fuel injectors. The degradation of each part is, with the current technology, impossible to detect without closer inspection. One example is fine cracks in turbine blades resulting from frequent temperature changes and centrifugal forces. Such cracks and their size can usually only be detected by a borescope or by removing the turbine blade completely and using X-ray imaging technology. It is essential to remember which types of degradation or failure modes can be detected in advance with prognostics and for which types regular maintenance intervals for thorough checks are still required (Department for Business Innovation and Skills, 2016; Shannon & Ackert, 2011). The NASA dataset can only be used to detect degradation and fault levels on the macroscopic level (modules) but not for individual components.

4. NASA CMAPSS Dataset

A NASA turbofan jet engine dataset is used, provided by the Prognostics Center of Excellence at NASA Ames (NASA, 2023). This dataset was previously generated using the Commercial Modular Aero-Propulsion System Simulation (CMAPSS), which was fed by recorded flight conditions onboard a commercial aircraft. The CMAPSS dataset is chosen for this study for several reasons: it is generally recognized to be the benchmark dataset for RUL prognostics and has thus been used for multiple studies. The findings can therefore be easily compared with existing research. Also, even the sub-dataset FD001 is extensive and contains over 20,000 cycles. A cycle in this context is defined as follows: when an engine undergoes engine start, takeoff, landing, and engine shutdown, this counts as one engine cycle.

4.1. Dataset Description

The CMAPSS dataset consists of 4 sub-datasets, labeled FD001 to FD004; each focused on different operating conditions and fault modes. The fault mode provides more details on how the engine degrades and ultimately fails, e.g., an HPC degradation or a Fan degradation are two of the occurring fault modes. The operating conditions are, e.g., the Mach number of the aircraft or the altitude.

To train the machine learning model and to later compare different maintenance strategies, only sub-dataset FD001 is considered, where the fault mode is an HPC degradation. Dataset FD001 contains data from 100 different engines. Each of those engines is of the same type, e.g., the PW1100 from Pratt & Whitney, but starts with a different degree of initial wear and manufacturing variations, resulting in varying cycles before the failure of each engine.

When planning the next engine inspection and determining the elapsed engine life, both the flight hours of the engine as well as the cycles are considered. This is because the high temperature within the engine and the high centrifugal forces of rotating components vary significantly during different flight phases. Long cruise phases where the engine is operated at relatively uniform thrust settings well below the maximum thrust setting deteriorate the engine life less significantly than high thrust settings during takeoff. If the aircraft is used for very short flights with relatively short cruise phases, the additional wear resulting from frequent load changes during the cycles would be underestimated if only the flight hours were considered. As this dataset does not provide any insights into the length of each flight in terms of flight hours, only the cycles are considered.

The dataset can be further divided into training and test datasets, each containing data from 100 engines. The training set contains sensor data of 20,631 cycles, while the test set contains sensor data of 13,096 cycles. In both datasets and for each cycle i, the unit number e_i of the engine, the current flight cycle c_i of engine e_i, multivariate time series sensor data (sensors s_{1i} to s_{21i}), as well as the operating conditions (o_{1i}, o_{2i}, o_{3i}) are provided. Therefore, each row with individual data points represents a snapshot taken during one cycle, representing one flight. Each sensor measures a specific engine parameter, such as the Total Temperature of the LPC outlet, Physical Core Speed, HPC Coolant Bleed, and Total Pressure at the High-Pressure Compressor outlet. The specific physical parameter measured by each sensor is listed in Table 20 in the appendix but shall not be discussed here in further detail. The dataset can be formally described as X:

$$\begin{gathered} X = [x_1, x_2, \ldots, x_N]; \\ \text{with } x_i = [e_i, c_i, s_{1i}, s_{2i}, \ldots, s_{21i}, o_{1i}, o_{2i}, o_{3i}] \\ \text{and } i \in \{1, 2, \ldots, N\} \end{gathered} \tag{1}$$

with N as the total number of flights, or cycles, in dataset FD001, and x_i containing all values provided during snapshot (or flight) i, which are the engine ID e_i, current cycle c_i, sensor data $s_{1i} \ldots s_{21i}$, and operating conditions o_{1i}, o_{2i}, o_{3i}. N has to be differentiated between the training data set ($N = 20,631$ cycles) and the test data ($N = 13,096$ cycles).

In the training set, the data of each engine starts at cycle 1, with each engine exhibiting an unknown engine history and different manufacturing variations. The HPC degrades from cycle to cycle until the engine eventually fails. In the test set, various cycles are extracted from engines during different phases of the engine life, which is unknown. The available test data of each engine ends abruptly with varying remaining cycles until engine failure; refer to Figure 2 for a graphical representation.

An additional file is provided containing the actual remaining cycles of each engine after the test data ends.

4.2. NASA CMAPSS Turbofan Data Set - Analysis and Data Preparation

The training dataset is analyzed and prepared in a primary step. This is necessary before Machine Learning algorithms can be applied. Furthermore, some relevant information regarding engine degradation behavior can be extracted by evaluating the trends of different sensors. This preliminary analysis can later help to validate the model.

In the training set, the total number of cycles of the 100 engines varies significantly between 120 and 360 cycles. The

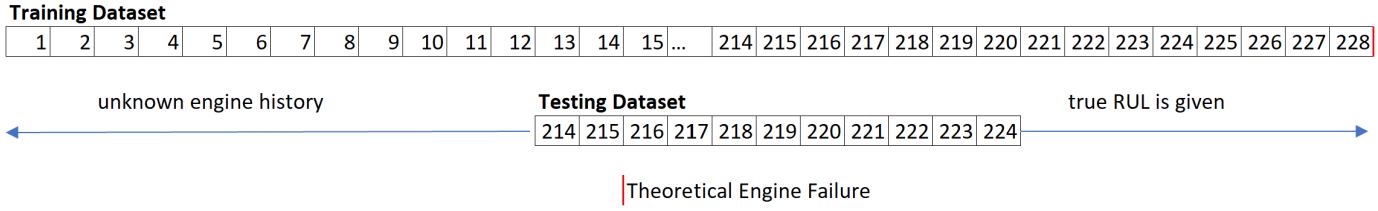


Figure 2: Graphical representation of the training and testing sets of the CMAPSS dataset

histograms of the 21 sensor values and the three operating conditions in Figure 3 show that sensors $s_1, s_5, s_{10}, s_{16}, s_{18}$, and s_{19}, and operating condition o_3, do not show any variability; those values remain constant throughout the entire dataset for all engines and all cycles. Since those six sensors and o_3 do not add any information for the machine learning model, they are removed from the dataset for the following steps. The values of the remaining sensor and operating conditions follow a bell curve, which may be highly skewed.

To visualize the trend of the remaining sensor values during engine operation, the values are plotted for ten randomly selected engines of the training set, see Figure 4. Each color represents one engine. Since each engine runs for a variable number of cycles before failure, the abscissa is reversed, so that cycle 0 (engine failure) is at the right for all engines to increase comparability. Some sensors show a clear trend, with sensor values either increasing or decreasing until failure (s_2, s_3, s_{11}). Some other sensors (e.g., s_{14}) also show a trend, but the values diverge during the final 75 cycles and either increase for some engines or decrease for others.

The range of each sensor is different; e.g., the s_2 values range from 640 to 645 while the s_{11} values range from 46 to 49. Feature rescaling is thus applied to each sensor and operating condition individually to normalize the range to the interval $[0, 1]$. To do this, the Min-Max Normalization is used, which can be described by eq. 2, with x' being the normalized sensor value and x being the original value:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2}$$

After normalization, the remaining sensors and operating conditions serve as the features for

the machine learning model, described as X' in eq. 3:

$$\begin{gathered} X' = \left[x'_1, x'_2, \ldots, x'_n\right]; \\ \text{with } x'_i = \left[s'_{2i}, \ldots, s'_{21i}, o'_{1i}, o'_{2i}\right] \text{ and } i \in \{1, 2, \ldots, N\} \end{gathered} \tag{3}$$

For each cycle i, the model output is the RUL, a single parameter. The RUL_i indicates the remaining cycles during cycle i until engine failure and serves as the label of the model Y. Formally, this can be described as:

$$\begin{gathered} Y = [y_1, y_2, \ldots, y_n]; \\ \text{with } y_i = [RUL_i] \text{ and } i \in \{1, 2, \ldots, N\} \end{gathered} \tag{4}$$

When examining the trend for each sensor in Figure 4 again, the values remain relatively constant until they start to show a distinct trend and either increase or decrease towards the end of their life. The closer the engine is to its failure, the more distinct this trend becomes, generally starting at approx. 100 cycles before failure. Machine Learning algorithms can extract only very little information during the phase where the sensor values are relatively constant. Several papers have thus concluded that the prediction performance of the RUL for cycles > 125 is low (dePater et al., 2022). Performance can be significantly improved by introducing a piecewise linear function as the new target function (label) during the training process, as depicted in Figure 5. If the target RUL is > 125, it is set to the constant value of 125, resulting in the new label Y':

$$\begin{gathered} Y' = \left[y'_1, y'_2, \ldots, y'_n\right]; \\ \text{with } y_i = \left[RUL'_i\right] \text{ and } i \in \{1, 2, \ldots, N\} \end{gathered} \tag{5}$$

5. Machine Learning Model

A primary analysis of different machine learning algorithms has been conducted to determine which algorithm exhibits the highest performance when applied to the CMAPSS dataset. During this study, five pre-selected algorithms were trained with the training data set and evaluated using the test data set. Criteria for the pre-selection process were: performance scores in existing literature, simplicity of implementation, and fitness to be applied to multivariate time-series data. Those algorithms are:

Decision Tree, Random Forest, K-Nearest Neighbor, Support Vector Machine, and MultiLayer Perceptron.

To evaluate the 5 described algorithms and different hyperparameters, the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE) are calculated and used as the model performance metrics, see eq. 6 and 7 (Trevisan, 2022):

$$\text{MAE} = \frac{\sum_{n=1}^{N} \sum_{m_n=1}^{M_n} |y_{mn} - x_{mn}|}{\sum_{n=1}^{N} M_n} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{n=1}^{N} \sum_{m_n=1}^{M_n} (y_{mn} - x_{mn})^2}{\sum_{n=1}^{N} M_n}} \tag{7}$$

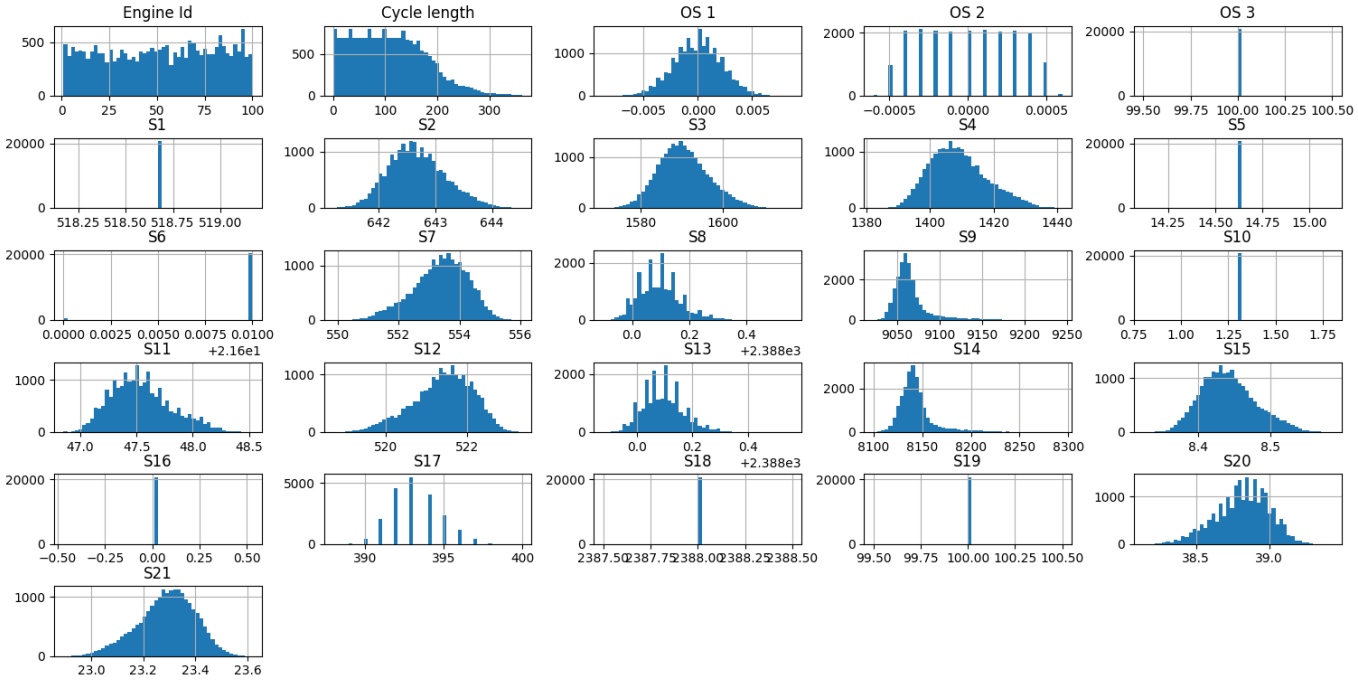


Figure 3: Histogram of CMAPSS training dataset, each window representing an individual sensor or operating condition

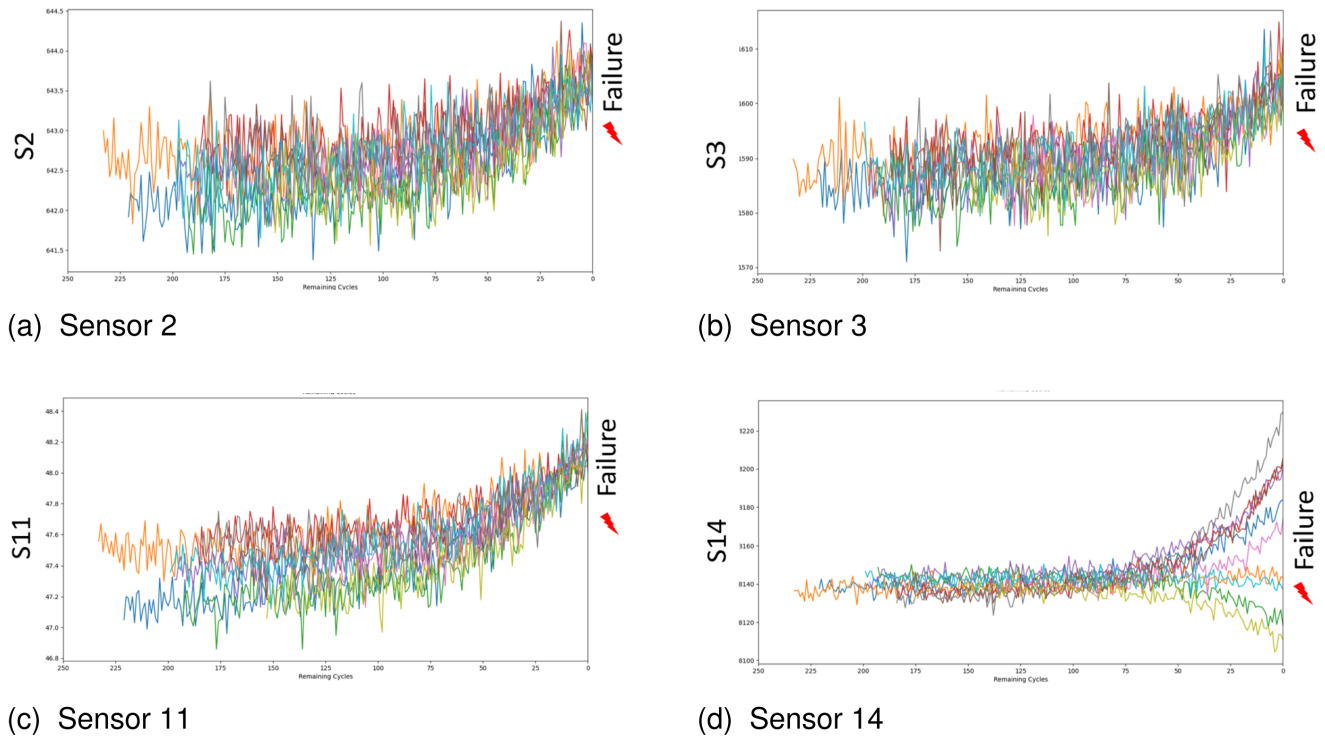


Figure 4: Values of ten randomly selected engines for four different sensors $(s_2, s_3, s_{11}, s_{14})$ from cycle 250 (left) to 0 (right)

With N: number of engines in the dataset, M_n: number of total cycles of engine n, y_{mn}: predicted RUL of engine n during cycle m, and x_{mn}: true RUL of engine n during cycle m.

5.1. Algorithms and Hyperparameter Tuning

For each algorithms, a preliminary hyperparameter tuning, sometimes referred to as hyperparameter optimization, has been conducted to optimize the performance. Hyperparameters are external model parameters that are not adapted

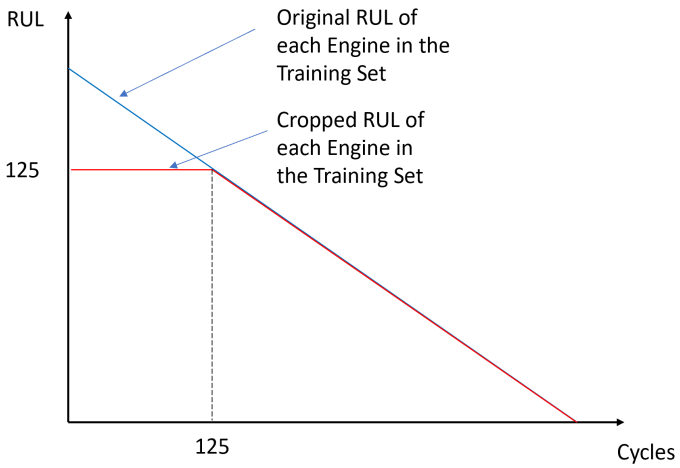


Figure 5: RUL Target Function, which serves as the label for the Machine Learning model

automatically during algorithm training but must be set manually to control the learning process. Therefore, hyperparameters differ from the system parameters, such as the weights of the Artificial Neural Network nodes, which are learned and optimized during the training process itself. Depending on the model, determining the optimal hyperparameters can be difficult, especially for complex applications and multivariate data. Hyperparameters must be balanced in the right way to prevent overfitting and underfitting. Therefore, they significantly influence the resulting model and must be determined as accurately as possible. The bounds for each hyperparameter are derived from literature and through experimentation, representing an appropriate balancing of computing power and model performance (An et al., 2015; scikit-learn developers, 2023; Silvestrin et al., 2019). There are different strategies for this hyperparameter tuning:

Grid Search is an exhaustive search where all combinations of the given hyperparameters are tested, and the optimal tuple of hyperparameters is chosen.

On the other hand, **Random Grid Search**, which is a variation of Grid Search, does not evaluate all hyperparameter combinations but only a certain number of random combinations. This is usually faster but can still outperform Grid Search, especially if the model only depends on a few hyperparameters. Since this thesis mainly focuses on the cost comparison of different maintenance strategies, more complex optimization algorithms such as Bayesian optimization, Gradient-based optimization, and Evolutionary optimization have not been considered here (scikit-learn developers, 2023).

Grid Search has been selected for determining the hyperparameters of the described algorithms, using the test dataset and cross-validation (CV) of 3. The performance (RMSE and MAE) of the hyperparameter tuple is then determined as the average of the three independent CV computations.

It must be noted that the determined hyperparameters are not necessarily globally optimal parameters. Further refinement of the search space would be required to increase the performance further. By comparing the algorithms' predictions after hyperparameter tuning with other results published in papers, it can be concluded that the achieved accuracy is very similar, and thus the hyperparameters are within an appropriate range. In the following, each of the considered algorithms, specific advantages, and disadvantages, as well as their hyperparameters, shall be explained briefly. For further details, please refer to the provided reference.

5.1.1. Decision Tree

Decision Trees are one of the most common and easy-to-implement algorithms. They can be used in various fields, such as classification, regression, and pattern identification, showing promising results. Their structure can be best described as a tree-like flow chart, consisting of decision nodes that split the dataset further, end nodes as the final clusters (leaves), and branches for connecting nodes, see Figure 6. During the learning process, the tree starts to construct itself by splitting the training dataset, which can be understood as the tree's root node (green). The emerging branches from the root node feed into the internal decision nodes (grey), which split the dataset further to form homogeneous splits based on specific decision rules. This process is repeated, and decision rules at the decision nodes are adapted until the data within each emerging node is similar enough. At this point, the predefined termination criterion prevents the further splitting of nodes, resulting in the final clusters or leaf nodes (blue/orange).

The classification rules are the paths from the tree's root to the individual leaves. The tree contains a finite number of end nodes, each representing one cluster for the final predictions. Using the decision tree to cluster the data into four categories, the decision tree would have four types of end nodes. If it is used for regression, the regression prediction is not smooth but rather piecewise constant. The more branches and leaves the tree has, the more clusters for regression are available and the fitter the model (Mohammed et al., 2017).

The advantages of this algorithm are that the trees can

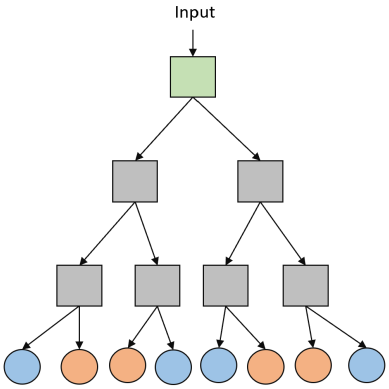


Figure 6: Graphical representation of a Decision Tree
(own figure, derived from Beauchamp (2020))

be visualized easily, and the branching process is thus easy to understand. There is also only very little data preparation required, such as normalization. On the other hand, very deep trees can be too complex for the underlying data, which means "overfitting" the training data can be a severe problem if no mechanisms to prevent this are in place (Mohammed et al., 2017). Since individual Decision Trees can be combined to form a Random Forest, the hyperparameters are only described for the Random Forest below.

5.1.2. Random Forest

Random Forests are multiple parallel individual Decision Trees. Each tree predicts the output from the input data as described above, but in a later step, the trees "vote" on the most common answer to improve the accuracy. This is graphically depicted in Figure 7. If the model is used for regression, the mean average of all individual trees is usually calculated and used as the final predicted value of the model. In case individual trees are overfitted or not trained well enough, the polling can improve the prediction of the forest significantly; therefore, they usually outperform individual trees (Mohammed et al., 2017). Table 1 provides the hyperparameters and their ranges (scikit-learn developers, 2023).

- **Number of Estimators:** number of individual trees within the forest
- **Max. Depth:** maximum depth of each tree
- **Min. Samples per Split:** minimum number of samples within each internal node required for a split.
- **Min. Samples per Leaf:** minimum number of samples required so that a node can become a leaf node; this value is directly linked to the "Min. Samples per Split", as they influence one another. Both hyperparameters are essential to smooth the model.
- **Criterion:** the function which is used to determine the quality of a split: Squared Error: evaluates new split by calculating the mean squared error; Friedmann MSE: calculates the squared error with Friedmann improvement. Poisson: uses the reduction in Poisson deviance.
- **Max. Features:** during each split, only one feature is considered.

As described for the Decision Tree, the structure of Random Forests with individual Decision Trees can be visualized easily. Another advantage is that the feature importance can be determined, which shows the influence each input feature (here: sensors and operating conditions) has on the prediction of the label (RUL). Details are provided in Table 2 for the five features with the highest and the lowest importance. When analyzing a Machine Learning model, the feature importance can help determine which specific sensors are highly important for the prediction of the output.

As Table 2 suggests, s_{11} shows the highest feature importance and accounts for roughly 64.67% of the prediction, followed by s_9(13.64%) and s_4 (6.72%). The five sensors with the highest feature importance explain more than 91% of the prediction, while the lowest-ranked sensors/operating conditions account for only 1.42% of the prediction.

5.1.3. K-Nearest Neighbors

K-Nearest Neighbors (KNN) algorithms can be used for supervised and unsupervised learning, using a proximity criterion to classify or predict the output of individual data points.

It can be used for classification and regression problems but is more commonly used for classification. Its implementation is simple and thus used for various applications, but the calculations can be computationally expensive if the number of data points is significant.

For classification, new data points are classified based on the majority vote of the surrounding k nearest datapoint neighbors. The most frequently represented label of the surrounding data points is then assigned to the new data point. Weights are commonly used, so closer data points contribute more strongly during the voting of the assigned class than more distant data points.

Table 1: Hyperparameters and the ranges for the Random Forest

Hyperparameter	Lower Bound	Upper Bound	Chosen Parameter
No. of Estimators	1	500	362
Max. Depth	1	100	9
Min. Samples for Split	2	20	7
Min. Samples per Leaf	2	20	17
Criterion	Squared Error; Friedmann MSE; Poisson		Poisson

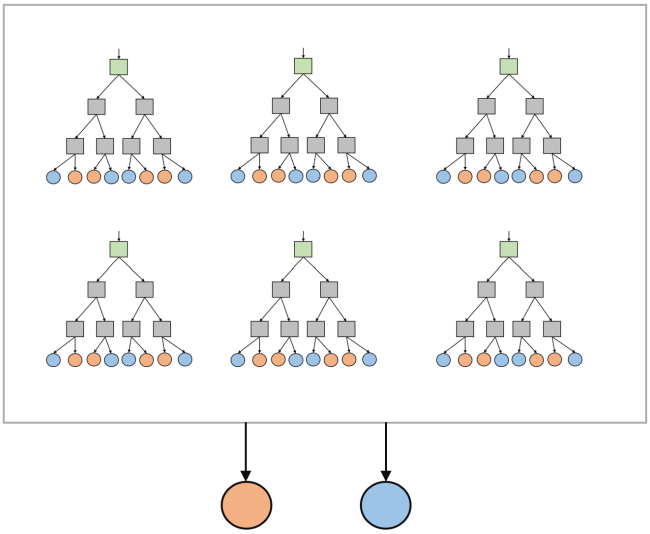


Figure 7: Graphical Representation of a Random Forest with individual Decision Trees (own figure, derived from Beauchamp (2020))

Table 2: Feature Importance for the Random Forest

Highest Importance	Feature Importance [%]	Lowest Importance	Feature Importance [%]
s_{11}	64.67	s_8	0.59
s_9	13.64	os_1	0.39
s_4	6.72	s_{17}	0.23
s_{12}	4.05	os_2	0.21
s_7	2.04	s_6	0.00
Sum:	**91.12**	**Sum:**	**1.41**

For regression tasks, the average value of the k nearest neighbors is assigned to the new data point. Again, weighting the votes of the surrounding data points based on their distance can improve the performance of the model.

How the distance is calculated between data points is essential and can significantly affect the model. Some distance measures commonly used are the Manhattan Distance and the Euclidean Distance. Also, the number of considered neighbors k has a significant impact on the performance of the algorithm (IBM, 2023). The hyperparameters and the ranges are provided in Table 3 (scikit-learn developers, 2023).

- **Number of Neighbors:** hyperparameter k specifies how many surrounding data points, or neighbors, are considered to determine the prediction output. This parameter has a significant impact on the model.
- **p**: parameter p represents the method to calculate the distance between data points. $p = 1$ is equivalent to using the Manhattan Distance, while $p = 2$ means the Euclidean Distance is used.
- **Weights:** "Uniform" allocates the same weight parameter to all data points during the calculations. For "Distance", the closer the data point, the higher the attributed weight parameter.

Table 3: Hyperparameters and the ranges for the KNN

Hyperparameter	Lower Bound	Upper Bound	Chosen Value
k	1	50	37
p	1	2	1
Weights	Uniform; Distance		Uniform

5.1.4. Artificial Neural Network

The Artificial Neural Network (ANN) can be described as a computing system that tries to mimic the biological nervous system.

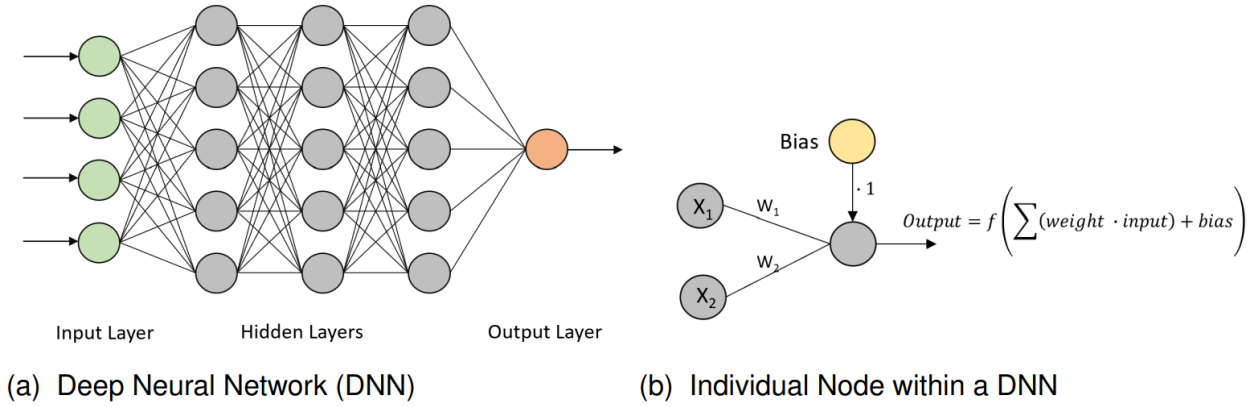
The ANN consists of a collection of artificial neurons, also called nodes, which are connected to transmit information via "edges". Each node can receive signals from the previous layer of nodes, then processes the input signals in a specific, non-linear way, which is represented by the activation function $f(x)$, and sends the result of the computation to the following layer. Each node and edge have an attributed weight parameter w to adjust the strength of each input signal. The higher the weight, the higher the influence of the input sent through an edge from one node to the next. This is graphically represented in Figure 8b. The bias is a constant term added to the computation within each neuron to shift the result to the positive or negative.

During training, the network adjusts the weight parameters automatically to adapt to the specific task and improve performance. Training data passes through the network, where each node processes the information, and the corresponding output is predicted. The error between the predicted output and the true output, which is provided in the training data as the label during supervised learning, is calculated. The weight parameters and bias are then adjusted to minimize this error. When further adjusting the weight parameters does not improve the error significantly, the learning process can be stopped (Mohammed et al., 2017; Silvestrin et al., 2019).

The arrangement of neurons is usually divided into layers, and each layer of nodes may perform a different mathematical transformation before passing the signal to the next layer. The first layer is referred to as the input layer, which consists of as many neurons as there are features in the data. The last layer of the network is the output layer, which transforms the values it receives from the neurons of the previous layers into the output. Each layer between the input and output layer is called a hidden layer. If the model consists of more than one hidden layer, it is referred to as a Deep Neural Network (DNN), see Figure 8a. The more layers the model has, the more complex its calculations can be. If the network is fully connected, meaning each neuron of one layer is connected via an edge to all neurons of the next layer, but without any edges to neurons of the same layer, it is called Multi-Layer Perceptron (MLP). In the following, the MLP form of the ANN will be considered (Mohammed et al., 2017).

MLPs can model non-linear processes, and they are thus used in a variety of different applications, including pattern recognition, medical diagnosis, data mining, and vehicle control. Since the input parameters, such as the number of nodes in each layer and the number of hidden layers, have a significant influence on the structure of the network and thus on the accuracy of the predictions, a good prior understanding of the problem as well as the network is required. Disadvantages of this type of network include requiring a relatively large amount of training data. Also, although ANNs and MLPs are very powerful across multiple applications, it is generally difficult to understand how the specific model functions, especially for DNNs with multiple hidden layers (Mohammed et al., 2017). Table 4 provides the hyperparameters that are considered in the tuning, as well as the upper and lower bound.

- **Hidden Layer Size:** this integer value specifies the number of hidden layers within the network, excluding the input and the output layer.
- **Neurons per Layer:** defines the number of neurons per network layer. Not all layers are required to have the same number of neurons, e.g. $(40, 50, 40)$ would specify a network with 40 neurons in the first hidden layer, 50 neurons in the second, and 40 in the third hidden layer.
- **Activation Function:** defines the specific function of each node within the hidden layers and how the nodes' input is transformed into the output. Two functions have proven to be extremely powerful and are thus considered in this hyperparameter tuning: tanh is defined as the hyperbolic tan function, which returns f(x) = tanh(x); and RELU (Rectified Linear Unit Function), which returns $f(x) = \max(0, x)$
- **Solver:** is used to determine the weights during the learning process. The three considered solvers are: adam: which is a stochastic gradient-based optimizer; sgd: an optimizer relying on stochastic gradient descent, and lbfgs: an optimizer that is part of the quasi-Newton methods.
- α: during the learning process, the error between the predicted output and the true output is minimized. In addition to this "plain error" term in the error function, a second term, the L2 term, is introduced, penalizing larger weight parameters to prevent overfitting. Larger values of α may prevent high variance, which is a sign of overfitting, while lower values of α encourage larger weights to prevent bias, a sign of underfitting.
- **Learning rate:** set to "adaptive", meaning it decreases during learning.



(a) Deep Neural Network (DNN) (b) Individual Node within a DNN

Figure 8: Graphical Representation of a Deep Neural Network (own figure, derived from McCaffrey (2013))

Table 4: Hyperparameters and the ranges for the MLP

Hyperparameter	Lower Bound	Upper Bound	Chosen Value
Hidden Layer Size	1	4	3
Neurons per Layer	5	100	(40, 40, 40)
Activation Function	Tanh; RELU		RELU
Solver	Adam; SGD; LBFGS		Adam
Alpha	0.00005	0.001	0.0002

5.1.5. Support Vector Machine

The Support Vector Machine (SVM) is a supervised algorithm that can be used for classification and regression tasks. For example, if a training dataset with data points belonging to one of two categories is provided, the algorithm builds a model by splitting the dataspace into two distinct areas.

During the learning phase, the split is optimized by maximizing the distance between the data points belonging to the two categories. By using a kernel trick to increase the dimension of the dataspace, the algorithm can also be used for non-linear classification. An extension to the described model allows regression, which is considered Support Vector Regression (SVR). SVM and SVR are used in various applications and provide good results. Generally, the performance is highest if there is a clear margin between the splits or if the dataspace can be easily described in a higher dimensional space. Since this concept is very complex, it is referred to Smola and Schoelkopf (2004) for further details and additional information. The result of the hyperparameter tuning is provided in Table 5.

- **C:** usually, it is not possible to split the dataset into distinct categories without some misclassified data points. Parameter C introduces a "soft margin" which allows the misclassification of some data points but penalizes each one.
- γ: when applying a Kernel trick, the parameter gamma controls the similarity measures in the new space.
- **Kernel:** RBF: Radial Basis Function Kernel; Poly: Polynomial Method.

5.1.6. Uncertainties

Regardless of the algorithm used to construct the model, the prediction is influenced by uncertainties. It is crucial to keep the sources of uncertainty in mind and reduce them as much as possible to improve the model's performance and, thus, the predictions. Generally, uncertainties can be categorized into (Huellermeier, 2021):

- **Epistemic Uncertainty:** this type of uncertainty is caused by a lack of relevant training data. Relevance in this context refers to data where the training and test data are similar. Depending on the type of problem, its complexity, and the chosen algorithm, a different amount of training data might be required. Generally, models have a low epistemic uncertainty if the training data and the test data are very similar, which means the model can extract the relevant information during the training process and then accurately predict the output when the test data is used. By providing additional relevant training data, this type of uncertainty is reducible (Huellermeier, 2021).
- The second source of uncertainty is called **aleatoric uncertainty**, which refers to the contamination of the data with noise or randomness. It is often referred to as data uncertainty. Unlike epistemic uncertainty, aleatoric uncertainty cannot be reduced by providing additional data (Huellermeier, 2021).

Table 5: Hyperparameters and the ranges for the SVR

Hyperparameter	Lower Bound	Upper Bound	Chosen Value
C	0.1	100	10
γ	0.0001	10	0.01
Kernel	RBF; Poly		RBF

5.2. Performance Analysis of Algorithms

The MAE and RMSE (eq. 6 and 7) are calculated for the five algorithms. As the results in Table 6 suggest, the Multi-Layer Perceptron provides the lowest MAE and RMSE, corresponding to the highest overall performance. Those values are within a similar range when other papers are considered (Silvestrin et al., 2019). The MLP is therefore chosen for the further procedure.

To evaluate the performance of the MLP more thoroughly, the difference between the true RUL and the predicted RUL is plotted for all cycles of the dataset as a histogram. Figure 9 shows the result, with the mean at 0.52, the mode at 0.62, and the red line at 0.0 for reference. The values to the right of the red line in the positive section of the abscissa can be considered uncritical because the predicted RUL values are lower than the actual RUL values (underestimated). All values to the left of the red line in the negative section of the abscissa are more critical because the MLP model overestimates the RUL.

When the predicted RUL (blue line) is plotted for an exemplary engine of the test set (engine 32) in Figure 10, the predicted RUL fluctuates due to the underlying background sensor noise of the data. To minimize the effect of those fluctuations, exponential smoothing is applied, which results in the red line. In eq. 8, s represents the smoothed data at cycle t, x the original, unsmoothed data, and the smoothing factor is set to $\alpha = 0.25$.

$$\begin{aligned} s_t &= \alpha \cdot x_t + (1-\alpha) \cdot x_{t-1};\ \text{for } t > 0 \\ s_0 &= x_0;\ \text{for } t = 0 \end{aligned} \tag{8}$$

Since this is an engine from the test set, the actual cycle at which the data recording begins is not known, but is set to 0 in the figure. The smoothed RUL remains relatively steady with minor fluctuations at around 125 for cycles > 80. This is the part of the piecewise linear function (label) where the target value is set to 125.

At around cycle 80, the predicted RUL starts to drop. The smoothed RUL ends cycle 144 with a predicted RUL of 54.6 cycles. The true RUL at this cycle is 48 cycles, meaning the MLP model overestimated the RUL by 6.6 cycles.

When different engines are compared, it is observed that the RUL for some engines is predicted more accurately than for other engines. To analyze this, the sensor values of "good" and "bad" predictions are compared in the same diagram. A "good" prediction in this context refers to engines where the difference between the last predicted RUL value and the actual RUL value differs < 4 cycles, while a "bad" prediction is defined to be a difference of > 25 cycles. Since the engines in the test set are all at different phases of engine life, only engines where the RUL after the last cycle is similar can be directly compared. In Figure 11, the sensor values of s_{11} and s_9 of three "good" engines (blue) and three "bad" engines (red) are plotted. The percentage value describes the feature importance of the respective sensor. For all sensors, there is no apparent difference between "good" predictions and "bad" predictions; therefore, some underlying information within the sensor data must be the reason for the prediction differences in the model.

6. Cost Comparison of Different Maintenance Strategies

In the following chapter, the considered maintenance strategies are compared based on the CMAPSS dataset. First, the methodology and the assumptions to determine the average maintenance costs per cycle $\bar{c}$ (see eq. 9) are introduced in Chapter 6.1. The general methodology is adapted when it is applied to each strategy in the subsequent sub-chapters. Also, different simplifications are implemented, as discussed below. The results of the cost comparison are provided in Chapter 6.2. The considered maintenance strategies are:

- **Reference:** $\bar{c}$ for the Reference Case shall serve as a lower baseline. For this, it is assumed that the true RUL of each engine is known in advance without error. Therefore, the required maintenance tasks can be conducted at the cycle during which engine failure would occur to maximize engine life. This reference case shall provide a theoretical lower cost limit per cycle, which cannot be replicated in a real scenario.
- **Reactive:** the engine always fails; no maintenance occurs before failure. This case shall serve as the upper limit for the costs per cycle $\bar{c}$.
- **Preventive:** the optimal maintenance interval t^* is determined based on the failure distribution of the engines. t^* is then applied to the respective dataset to determine $\bar{c}$.
- **Predictive:** the RUL of each engine is continuously determined using the described MLP model of Chapter 5.1 and 5.2. If the RUL falls below an alarm trigger T, the maintenance task is scheduled.

6.1. Methodology and Assumptions

To make engine maintenance costs comparable, they are commonly specified in terms of Costs / Flight Hours (FH) (Shannon & Ackert, 2011). Since the CMAPSS dataset lacks

Table 6: Performance comparison for different algorithms for the RUL predictions before and after Exponential Smoothing (Exp. Sm.)

Algorithm	Exp. Sm. MAE	Exp. Sm. RMSE	MAE	RMSE
Decision Tree	12.10	16.31	12.60	17.70
Random Forest	12.05	16.18	12.14	16.55
KNN	12.81	16.84	12.79	17.12
MLP	**10.45**	**15.94**	**10.65**	**16.08**
SVR	13.31	16.58	13.52	17.26

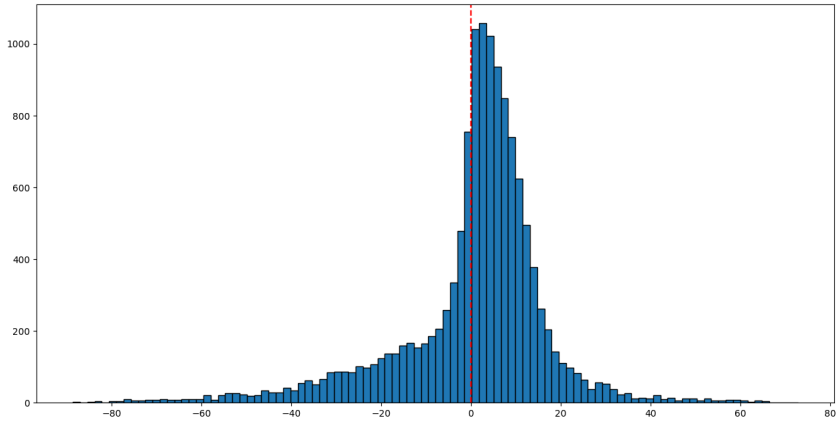


Figure 9: Histogram of Prediction Error

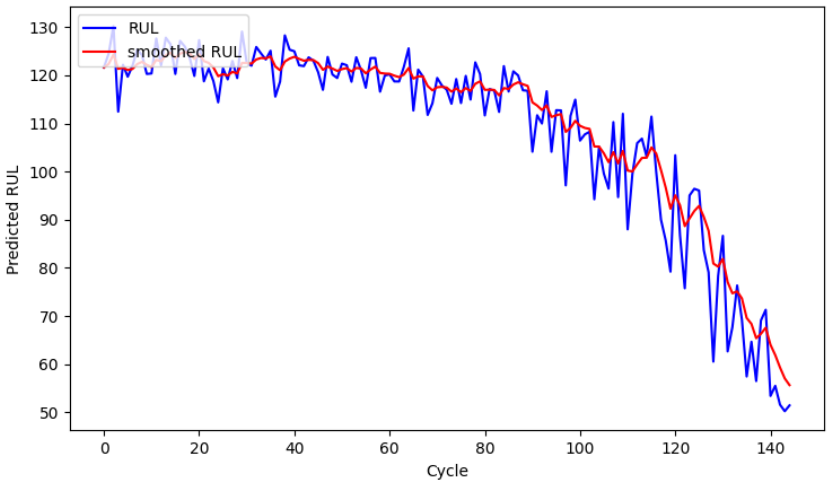


Figure 10: RUL Prediction of Engine 32 (blue: un-smoothed; red: smoothed)

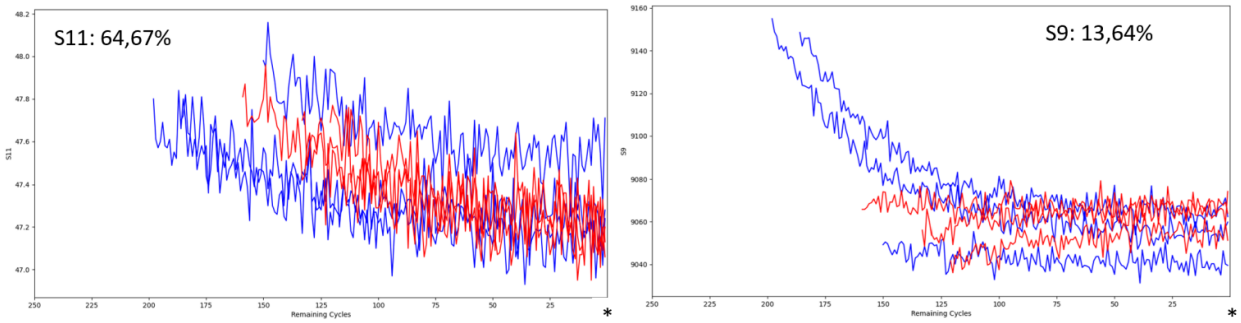
information about the FH, a different approach is chosen. Therefore, the average maintenance costs per cycle $\bar{c}$ are defined as:

$$\bar{c} = \frac{1}{N} \sum_{n=1}^{N} \frac{C_n}{Z_n} \tag{9}$$

with N: total number of engines in the considered sub-dataset, n: current engine, C_n the total maintenance costs for engine n, Z_n the total cycles of engine n.

The exact calculation of the average costs per cycle $\bar{c}$ as defined in eq. 9 differs slightly for the individual maintenance strategies. Therefore, an adaption of eq. 9 is required, as described in more detail below.

As the exact maintenance costs vary significantly based on the specific engine type and the conducted maintenance task, exemplary costs will be assumed for the following comparison. Artificial "Monetary Units" (MU) are used to show the relation of costs between different strategies rather than absolute values. The assumed cost to perform a maintenance task before failure is set to $c_m = 10$ MU. As a simplification,



(a) Sensor 11 values for different engines (b) Sensor 9 values for different engines

Figure 11: Sensor Values of s_{11} and s_9 for three "good Predictions" (blue) and three "bad" predictions (red). The percentage indicates the feature importance as determined by the Random Forest.

c_m does not vary irrespective of the maintenance task cycle (i.e., the costs are always 10 MU, regardless of whether 100 cycles remain until failure or just one cycle).

On the other hand, the costs to perform a maintenance task after failure are set to $c_f = 100$ MU, which is ten times higher than c_m. After the engine has failed, the tasks required to restore the engine are more complex due to internal damage and, thus, more costly. Also, if an aircraft engine fails during flight, this will usually result in an emergency landing at the closest airport, requiring additional costs to transport the failed engine to the next suitable maintenance facility. Unexpected aircraft downtime and negative media coverage of the incident may also decrease passenger volume and revenue, in the latter case, even for an extended time. Generally, these costs are complicated to estimate, but they shall be reflected in c_f.

It is assumed that at the first cycle of each engine, the engine is in the "as good as new" state, and both maintenance cases (maintenance before or after failure) will restore the engine to the "as good as new" condition again. Furthermore, for simplicity, each cycle shall represent one day.

To determine the costs per cycle and to compare the different maintenance strategies, engine data with RUL prediction until failure is required. For each engine of the test set, the data begins at an unknown cycle during the engine life and terminates some time before failure, as depicted in Figure 2. Determining the costs per cycle according to eq. 9 is thus not feasible because the total number of cycles is unknown. Therefore, the training set must be considered for the further analysis.

Since this training set is also necessary to train the Machine Learning model for Predictive Maintenance, the original training set is split five times to create new training and testing data. First, engines 1-20 are set aside and used as a new test set (TS1), while the remaining 80 engines (21−100) are used for model training (or learning), from now on referred to as learning set 1, or LS1. Next, engines $21-40$ are used as the new test set TS2, while engines 1-20 and 41-100 are used to train the model (LS2). This is repeated five times; the resulting new test sets are labeled TS1 to TS5, while the new learning sets are labeled LS1 to LS5. Although training instances are only necessary for Predictive Maintenance, the costs for all considered maintenance strategies are determined based on the same dataset to enable a comparison.

6.1.1. Reference Case

For this case, it is assumed that the true RUL of each cycle is known; therefore, the exact failure point (red line in Figure 12, representing a potential failure during cycle 228) can be anticipated well in advance. The optimal maintenance time (green) is the cycle during which engine failure would occur, which in this exemplary case corresponds to cycle 228. Eq. 9 is adjusted accordingly to determine the average costs per cycle of the Reference Case $\bar{c}_{ref}$, which results in:

$$\bar{c}_{ref} = \frac{1}{N} \sum_{n=1}^{N} \frac{c_m}{Z_n - 1} \tag{10}$$

It must be noted that this Reference Case can only be included in this analysis because the actual RUL values are provided in the dataset. In practice, this information is unavailable; therefore, such a reference case may not be feasible. Again, this shall serve as a lower baseline.

6.1.2. Reactive Maintenance

For this case, it is assumed that the engine always runs until failure and is then restored to the "as good as new" state. As described above, the incurred maintenance costs for maintenance after failure c_f are ten times higher than before failure c_m, which results in significantly higher average costs per cycle for Reactive Maintenance $\bar{c}_R$. Eq. 9 is again adjusted, leading to eq. 11.

This is graphically represented in Figure 13, where engine failure occurs during cycle 228.

$$\bar{c}_R = \frac{1}{N} \sum_{n=1}^{N} \frac{c_f}{Z_n} \tag{11}$$

Due to regulatory requirements, it is unlikely that no engine inspection or maintenance tasks are conducted during

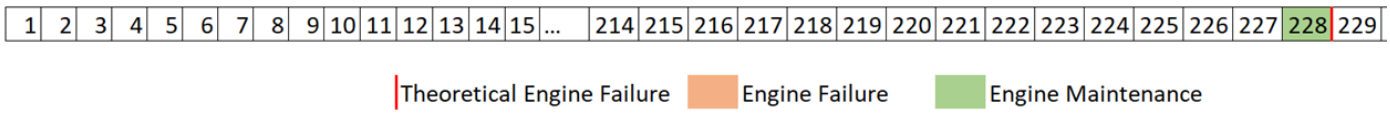


Figure 12: Graphical Representation of the Reference Case

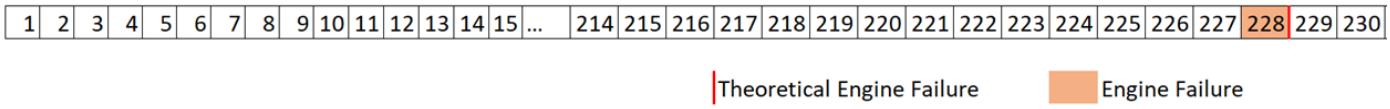


Figure 13: Graphical Representation of Reactive Maintenance

operation at all. Therefore, $\bar{c}_R$ shall serve as the hypothetical upper-cost limit.

6.1.3. Preventive Maintenance

As described in Chapter 3, the optimal maintenance interval for Preventive Maintenance is determined based on the failure distribution of the machine, or in this case, the engine. The approach is described for LS1 as an example, which is then repeated for LS2 to LS5.

In the primary step, the most suitable distribution function based on the Bayesian Information Criterion (BIC) and the Anderson-Darling Test (AD) is determined for LS1. The BIC is a commonly used criterion for distribution model selection. A distribution function with a low BIC value is usually preferred. The Anderson-Darling Test is a statistical test that helps to determine if a sample was drawn from data with a specific distribution. Again, a lower AD value generally indicates a better fit. For the detailed calculation of both test values, please refer to National Institute of Standards and Technology (2022) and Reid (2023).

The failure distribution of the 80 engines of LS1, as well as the probability density and the cumulative distribution of all fitted distribution functions, is shown in Figure 14. The results of the statistical tests for three common distributions are provided in Table 7. The probability plots of all fitted distributions is provided in the appendix in Figure 26 and Figure 27.

Based on the BIC and AD values in Table 7, the two-parameter Lognormal distribution describes the dataset most accurately. Therefore, the hypothesis that LS1 can be described by the Lognormal distribution is assumed. To test whether this hypothesis can be accepted or rejected, a Kolmogorov-Smirnov Test (KS) was conducted. This statistical test can determine the goodness of fit between the dataset and the chosen distribution, but it differs from the BIC and AD as it cannot be used for distribution comparison (National Institute of Standards and Technology, 2022; Reid, 2023). The KS test results prove that at the 0.05 significance level, the hypothesis that the dataset can be described by the lognormal distribution can be accepted (KS statistic value: 0.09412 for $\mu = 5.3062$ and $\sigma = 0.2121$).

On the other hand, the hypothesis that the dataset can be described by either the Normal distribution or the Weibull distribution can be rejected at the 0.05 significance level (KS statistic values of 0.1383 and 0.1566, respectively). Therefore, the Lognormal distribution is chosen to determine the optimal maintenance interval t^* for LS1. Statistical tests on all other learning sets LS2 to LS5 indicate that the Lognormal distribution also describes the individual failure distribution most accurately, but with different parameters μ and σ.

The Lognormal distribution can be described as a continuous probability distribution function of a variable whose logarithm is normally distributed (Dodge, 2008). Compared to the normal distribution, the Lognormal distribution is right-tailed. It is often used to model the distribution of technical or biological processes, such as system failures, where the variable cannot be negative. The Lognormal distribution is also commonly used in reliability analysis to model the time to perform Maintenance on a system, as is the case here (Dodge, 2008).

For LS1, the two required parameters describing the distribution are $\mu = 5.3062$ and $\sigma = 0.2121$. This results in a reliability function:

$$R(t) = \int_{\ln(t)}^{\infty} \frac{1}{\sigma' \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu'}{\sigma'}\right)^2} dx \tag{12}$$

with μ' as the mean of the natural logarithms of the times-to-failure and σ' as the standard deviation of the natural logarithms of the times-to-failure (Dodge, 2008).

As is the case for the Normal distribution, $R(t)$ for the Lognormal distribution does not have a closed form solution, meaning the corresponding value must be determined through a table or a numerical approximation. A polynomial of degree six is used for the numerical approximation of $R(t)$, from now on referred to as $R'(t)$, which shall not be explained further.

To determine the optimal maintenance interval t^*, the Cost Per Unit Time (CPUT) function is introduced in eq. 13:

$$CPUT(t) = \frac{c_m \cdot R'(t) + c_f \cdot (1 - R'(t))}{\int_0^t R'(s) ds} \tag{13}$$

with $R'(t)$ as the approximated Reliability Function at cycle t, and c_m and c_f as the maintenance costs before and after

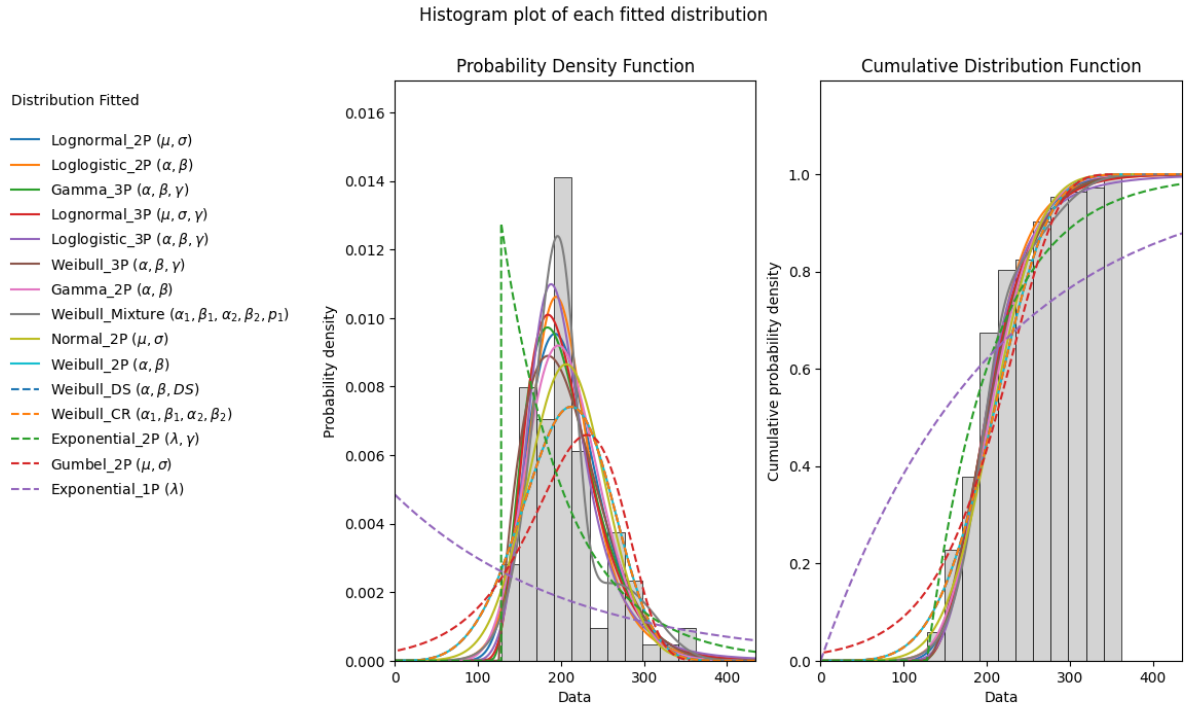


Figure 14: Probability Plot for all Distributions

Table 7: Statistical test results for three common distributions that are applied to LS1

Distribution	BIC	AD	KS
Lognormal 2P	**1044.12**	**0.8206**	**0,09412**
Weibull 2P	1059.21	2.0809	0,1383
Normal 2P	1070.71	3.0891	0,1566

failure, respectively (ReliaSoft Corporation, 2009). The optimal maintenance interval t^* is obtained by setting the partial differential equal to 0 (ReliaSoft Corporation, 2009):

$$\frac{\partial CPUT(t^*)}{\partial t} = 0 \tag{14}$$

After t^* is determined from the LS according to eq. 13 and eq. 14, the average costs per cycle $\bar{c}_{Pv}$ are then calculated for the respective TS as:

$$\bar{c}_{Pv} = \frac{1}{N}\sum_{n=1}^{N}\frac{c_m \cdot x_n + c_f \cdot (1-x_n)}{Z_n} \tag{15}$$

x_i can take the values of 0 and 1,0 meaning the engine has failed before the scheduled maintenance task was conducted, and 1 meaning the task has been conducted before failure. Z_n in this formula represents the cycle at which the maintenance task of engine n is conducted. If a failure occurs, Z_n is equal to the failure cycle of engine n, otherwise Z_n is equal to the optimal maintenance interval t^*. Results are provided in Table 10.

6.1.4. Predictive Maintenance

Lastly, the average maintenance costs per cycle for Predictive Maintenance $\bar{c}_{Pd}$ are determined. For each of the five dataset splits described above, the 80 engines of the respective LS are used to train an individual MLP model, while the remaining 20 engines of the corresponding TS are used to predict the RUL.

The RUL predictions of the 20 engines of TS1 are displayed in Figure 15. Each color represents a different engine, while the linear red line shows the true RUL at each cycle. Due to the previously described data noise, the predicted RUL is smoothed with exponential smoothing ($\alpha = 0,25$) according to eq. 8. This decreases the the fluctuations of the predictions, see Figure 16.

As Figure 16 suggests, the RUL prediction of the MLP model is overestimated for some engines and underestimated for others. Towards the end of the engine life, the RUL of all engines converges, and the prediction error (difference between the red line as the true RUL and the predictions) decreases. This can be confirmed when the average MAE across all 20 engines is plotted for cycles 125 to 0, see Figure 17. A decrease of the MAE can be observed from cycle 60 to 0, meaning the RUL prediction error decreases as the engine ap-

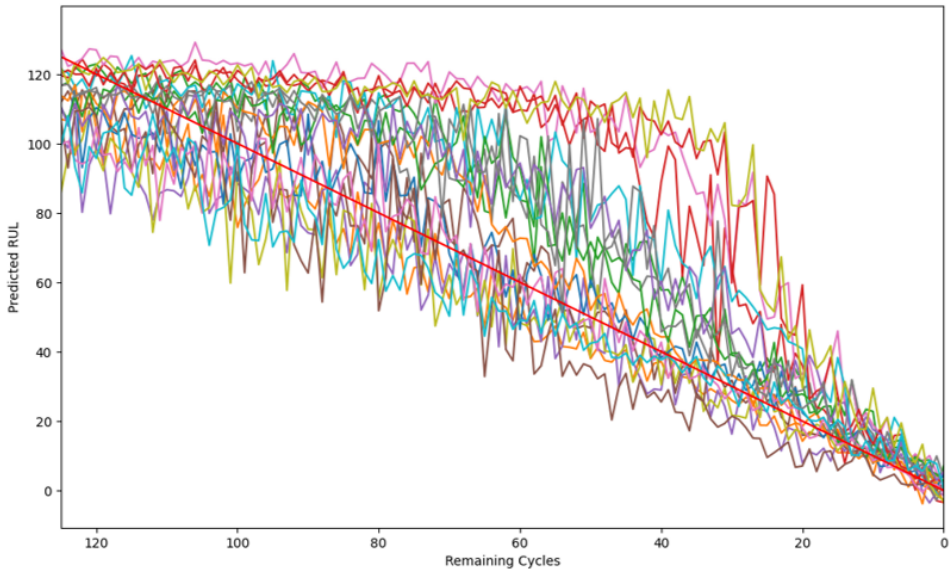


Figure 15: RUL predictions of 20 engines of TS1 before Exponential Smoothing

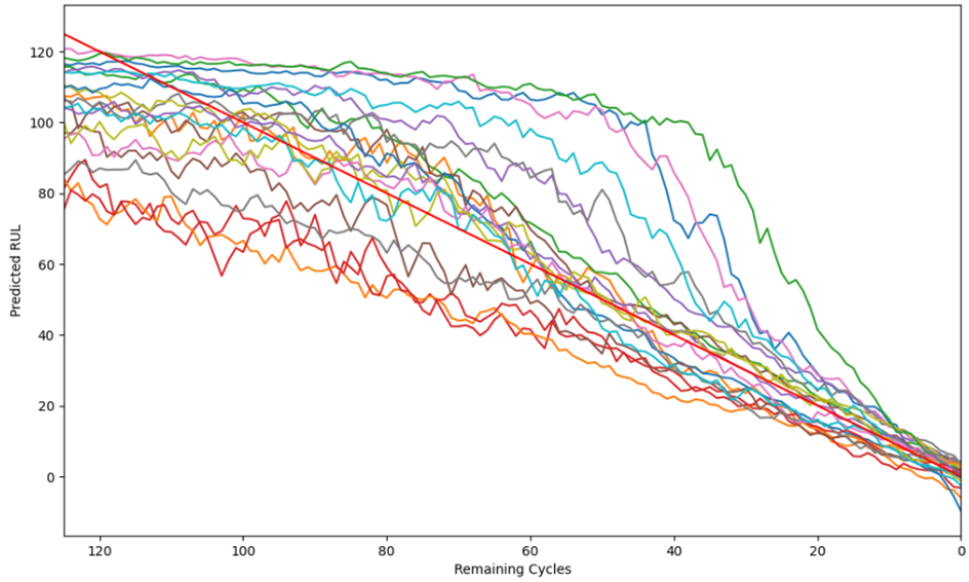


Figure 16: RUL predictions of 20 engines of TS1 after Exponential Smoothing

proaches engine failure. The same behavior can be observed for the RMSE (Figure 28 in the appendix).

As opposed to the other maintenance strategies, Predictive Maintenance considers the actual RUL of the respective engine, which is continuously predicted during each cycle. To determine how the maintenance decision process is influenced by the predicted RUL, an alarm trigger T and an operational buffer β are introduced.

As soon as the exponentially smoothed RUL of an engine falls below the alarm trigger T, a maintenance task is scheduled in exactly $\beta = 20$ days, corresponding to the operational buffer time. β is assumed to be required to prepare the engine maintenance task, i.e. to ensure sufficient time is available for spare parts procurement and capacity planning. Furthermore, airlines try to keep each aircraft in the air as much as possible and apply tight schedules without large buffers to incorporate delays. Therefore, enough time is required to reschedule the flights during the maintenance task to a different aircraft. β is a fixed parameter that is provided by external circumstances. It can only be influenced (decreased) by the maintenance operations but not by adapting the MLP model.

This concept is also depicted in Figure 18 for an exemplary alarm trigger of $T = 54$. The additional parameter γ shall represent the safety buffer, i.e., the remaining cycles after the scheduled maintenance task. Only T and β are speci-

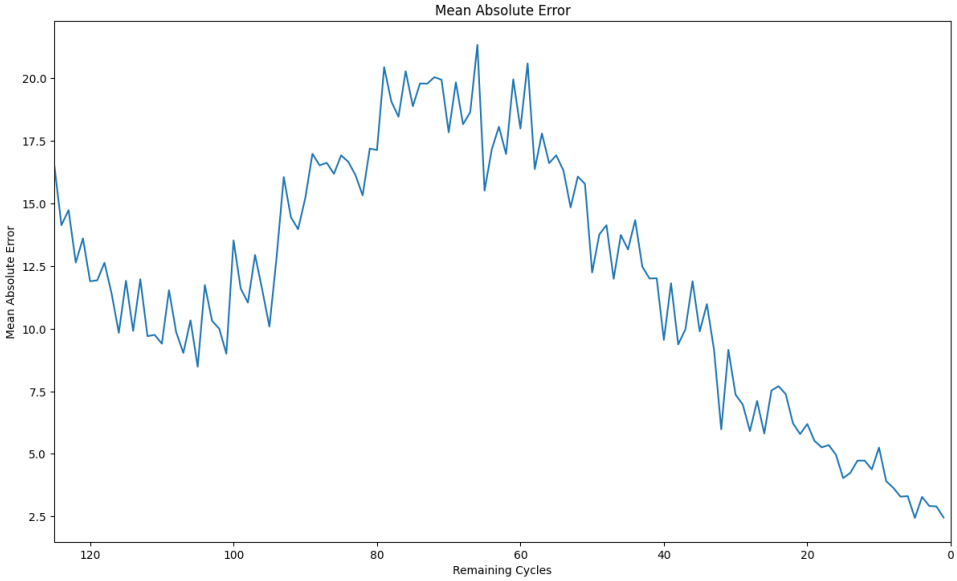


Figure 17: MAE for the engines of TS1

fied, while γ results from those two values automatically and can be calculated as $\gamma = T - \beta$.

The costs per cycle for Predictive Maintenance $\bar{c}_{Pd}$ are determined according to eq. 16:

$$\bar{c}_{Pd} = \frac{1}{N} \sum_{n=1}^{N} \frac{c_m \cdot x_n + c_f \cdot (1 - x_n)}{Z_n} \tag{16}$$

As previously described for Preventive Maintenance, x_i can take the values of 0 and 1, depending if failure occurs. Z_n corresponds to the cycle at which the maintenance task of engine n is conducted. If a failure occurs, Z_n is equal to the failure cycle of engine n.

To determine the optimal alarm trigger T^* for each TS, $\bar{c}_{Pd}$ is calculated according to eq. 16 for $T \in [20, 120]$. To maximize engine availability, T would ideally be equal the operational buffer β so that the engine is maintained as close to failure as possible. This would result in a safety buffer $\gamma = 0$. Due to the fluctuations and the uncertainty of the predicted RUL (primarily due to overpredicted RUL for some engines), if T is set too low, some engine would fail before the scheduled task in 20 days occurs. Therefore, T must be adjusted upwards until a cost-per-cycle minimum is reached.

As stated above, the optimal alarm trigger T^* is determined directly from the test sets. Generally, the preferred approach would be to use three different datasets: first, the machine learning model is trained with the engines of the training datasets (LS). Then, the resulting model is used to predict the RUL of different engines to determine T^*. Lastly, an independent third set of engines is used for the calculation of $\bar{c}_{Pd}$ with the previously determined T^*. However, this would require a three-fold split of the dataset, such as 60/20/20. Since only 100 training engines are available in the CMAPSS dataset, this approach is discarded because it would result in a machine learning model with lower performance. It is important to note that T^* can only be determined from each of the five test sets (TS1 to TS5) but not from the learning sets (LS1 to LS5). It is impossible to determine the performance of a machine learning model with the same data already used during the training process.

6.2. Results and Discussion for the Cost Comparison

The results of each maintenance strategy are provided in the following tables. Each TS is considered individually. The average maintenance costs per cycle $\bar{c}$, as well as the failure count (Failures) and the average interval to the maintenance task (Avg. Int.) are determined. If not specified otherwise, $c_m = 10$ and $c_f = 100$ are assumed.

6.2.1. Reference Case

The average costs per cycle for the Reference Case $\bar{c}_{\text{ref}}$ are summarized in Table 8, averaging at 0.05097 MU per cycle. Fluctuation can be observed across the Test Sets. The average interval (Avg. Int.) is determined to be 205.8 cycles. Since all engines are maintained before failure, the failure count (Failures) is zero for all Test Sets.

As this case serves as the lower baseline, the percentage increase of $\bar{c}$ of the following strategies related to the Reference Case shall also be determined.

6.2.2. Reactive Maintenance

For Reactive Maintenance, the results are summarized in Table 9. The average maintenance costs per cycle $\bar{c}_R$ across all TS is determined to be 0.5070 MU, while the average interval is 206.8 cycles. Compared to the Reference Case, the percentage increase of $\bar{c}_R$ (Incr. to Ref.) is 894.7%. Since no maintenance tasks are conducted, all 20 engines fail. Again, this case shall serve as the upper cost limit.

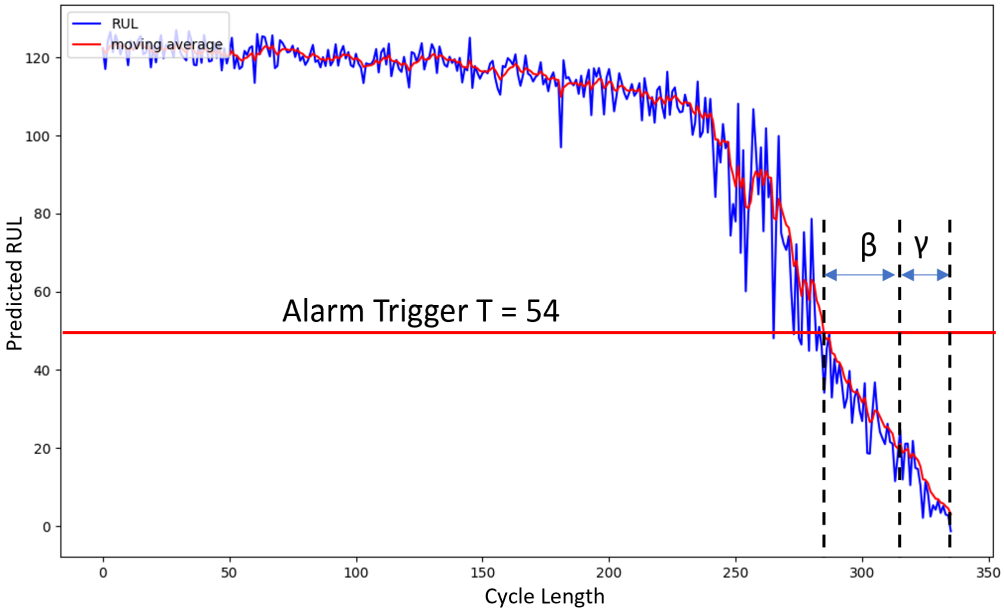


Figure 18: Graphical representation of the proposed methodology for Predictive Maintenance

Table 8: Cost comparison results for the Reference Case

Test Set	$\bar{c}_{\text{ref}}$	Failures	Avg. Int.
TS1	0.04992	0	207.4
TS2	0.05613	0	181.9
TS3	0.05035	0	208.4
TS4	0.05057	0	208.8
TS5	0.04788	0	223.7
Average	0.05097	0	205.8

Table 9: Cost comparison results for Reactive Maintenance

Test Set	$\bar{c}_R$	Failures	Avg. Int.	Incr. to Ref. [%]
TS1	0.4966	20	208.4	894.8
TS2	0.5581	20	182.9	894.3
TS3	0.5009	20	209.4	894.8
TS4	0.5031	20	209.8	894.9
TS5	0.4764	20	224.7	895.0
Average	0.5070	20	206.8	894.7

6.2.3. Preventive Maintenance

For the individual distribution parameters of μ and σ (depending on the respective TS), the optimal maintenance interval is determined to be $t^* = 122.9$ cycles (average across all TS), according to eq. 13 and eq. 14. Compared to the Reference Case, this a drastic decrease. t^* is rounded down to obtain integer values, because decimal intervals are not practicable. According to eq. 15, the optimal average costs per cycle for Preventive Maintenance can be determined to be $\bar{c}_{Pv} = 0.0885$ MU, which is a 73.6% increase compared to the Reference Case. Across all TS, no engine failures occur.

The two terms of the numerator of $CPUT(t)$ (eq. 13) for TS1 are depicted in Figure 19 in grey (Corrective Repair Costs) and orange (Preventive Repair Costs) for different maintenance intervals t. CPUT (t) is depicted in the same figure in blue (approximated by a polynomial function, deg. 6)

As can be seen, the minimum is located at $t^* = 119.90$ cycles, which is the same result already obtained by eq. 14. If a maintenance interval $t > t^*$ is chosen, $\bar{c}_{Pv}$ increases due to an increased chance of engine failure. On the other hand, if a maintenance interval $t < t^*$ is chosen, $\bar{c}_{Pv}$ increases due to over-maintenance.

Furthermore, the optimal maintenance interval t^* depends on the cost assumption of c_f and c_m, see the eq. 13. To determine the influence of varying cost assumptions on t^* and $\bar{c}_{Pv}$, different cost ratios of c_f/c_m shall be considered for

Table 10: Cost comparison results for Preventive Maintenance

Test Set	t^*	$\bar{c}_{Pv}$	Failures	Incr. to Ref. [%]
TS1	119.9	0.0902	0	80.6
TS2	124.7	0.0875	0	55.8
TS3	120.5	0.0905	0	79.7
TS4	124.3	0.0878	0	73.6
TS5	125.2	0.0864	0	80.5
Average	122.9	0.0885	0	73.6

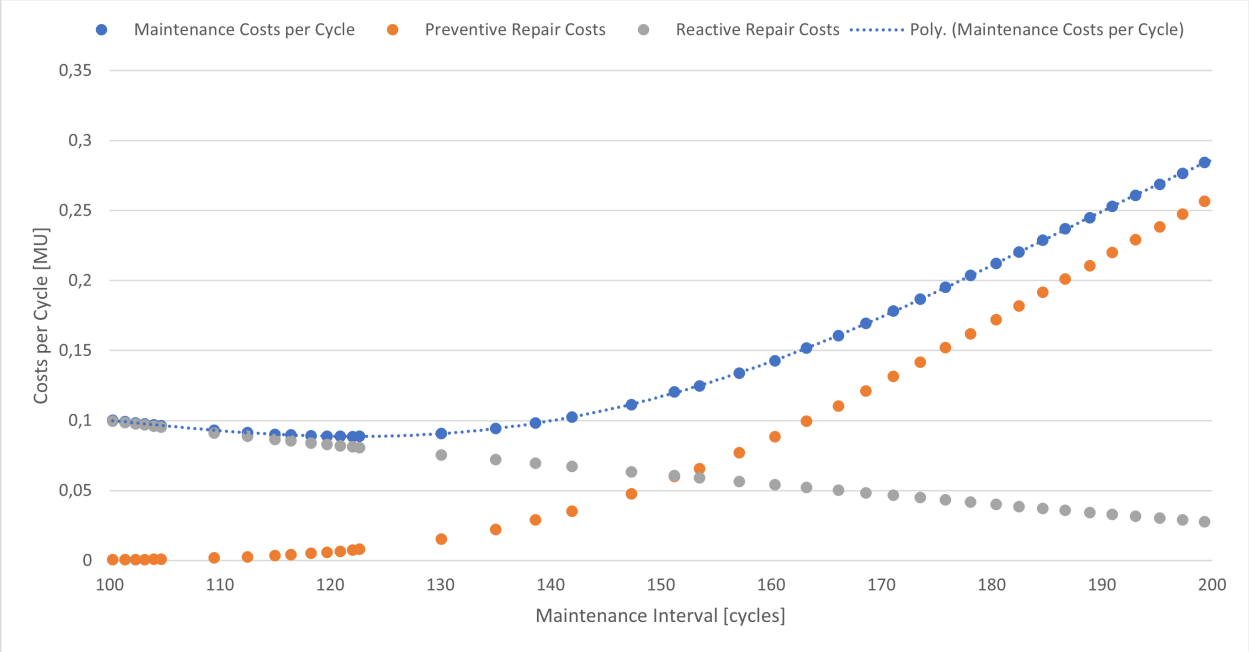


Figure 19: Preventive Maintenance Costs

TS1. Therefore, c_f is varied between $c_f = [20, 1000]$, while $c_m = 10$ is kept constant.

Table 11 provides t^* and the resulting $\bar{c}_{Pv}$ for the different cost ratios. This analysis shows that $\bar{c}_{Pv}$ increase as the ratio of c_f/c_m increases. Also, for higher ratios of c_f/c_m, the optimal maintenance interval shifts to shorter intervals because it is more favorable to decrease the chance of failure. Those results are also graphically depicted in Figure 20 a and b.

6.2.4. Predictive Maintenance

The optimal T^* for each Test Set and the resulting costs per cycle $\bar{c}_{Pd}$ according to eq.16 are provided in Table 12:

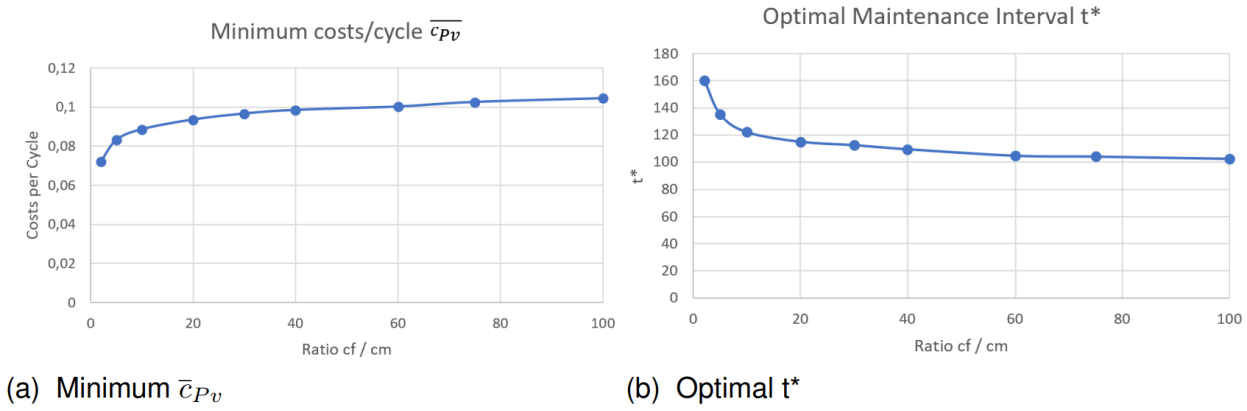
When $\bar{c}_{Pd}$ between the Reference Case and Predictive Maintenance is compared, a cost increase of 13.4% compared to the Reference Case is observed. On the other hand, when Preventive Maintenance and Predictive Maintenance are compared, $\bar{c}_{Pd}$ is on average 30.1% lower compared to $\bar{c}_{Pv}$. Also, the average interval for Predictive Maintenance is 58.1 cycles longer compared to Preventive Maintenance. Therefore, a drastic cost decrease can be realized by employing Predictive Maintenance.

The optimal Alarm Trigger T^* and the resulting average costs per cycle $\bar{c}_{Pd}$ strongly depend on the assumption of the operational buffer β. To demonstrate this influence, β is varied within the interval $\beta = [1, 50]$ for TS1. For each β, the optimal Alarm Trigger T^* and the costs per cycle $\bar{c}_{Pd}$ are determined individually, as described in Chapter 6.1.4.

Table 13 shows that lower operational buffers β result in lower lower optimal Alarm Triggers and lower costs per cycle. As shown in Figure 17, the MAE decreases as the engines approach engine failure. Therefore, a low Alarm Trigger close to failure is preferred to decrease $\bar{c}_{Pd}$. If $\beta = 0$ were chosen, an Alarm Trigger of $T^* = 4$ would be sufficient, and $\bar{c}_{Pd}$ would only be 3.2% higher than the Reference Case. Even with a rather pessimistically constraint of the chosen operational buffer $\beta = 20$ days, Predictive Maintenance outperforms Reactive and Preventive Maintenance.

7. Maintenance Framework for a Fleet of Aircraft comparison of different maintenance strategies

The cost comparison methodology of the previous chapter shall be applied to a more realistic scenario. Therefore, a maintenance planning framework to determine and compare the average maintenance costs per cycle $\bar{c}$ is proposed.



(a) Minimum $\bar{c}_{Pv}$ (b) Optimal t*

Figure 20: Graphical representation of the relation of varying cost ratios and $\bar{c}_{Pv}$ and t*

Table 11: Varying Cost Ratios

c_f/c_m	c_m	c_f	t^*	$\bar{c}_{Pv}$
2	10	20	160.35	0.07201
5	10	50	135.02	0.08312
10	10	100	119.90	0.09020
20	10	200	114.99	0.09359
30	10	300	112.49	0.09665
40	10	400	109.46	0.09849
60	10	600	104,66	0.10023
75	10	750	104.00	0.10256
100	10	1000	102.37	0.10446

Table 12: Cost comparison results for Predictive Maintenance

Test Set	T^*	$\bar{c}_{Pd}$	Failures	Avg. Int.	Incr. to Ref. [%]
TS1	41	0.05491	0	189.4	10.0
TS2	54	0.06933	0	162.6	12.1
TS3	51	0.05975	0	183.3	10.6
TS4	64	0.06572	0	173.7	20.6
TS5	57	0.05887	0	195.8	13.7
Average	53	0.06187	0	181.0	13.4

Table 13: Results for varying operational buffers β for Predictive Maintenance

β	T^*	$\bar{c}_{Pd}$	Failures
0	4	0.05153	0
1	6	0.05177	0
5	15	0.05235	0
10	22	0.05293	0
15	28	0.05322	0
20	41	0.05491	0
25	56	0.05911	0
30	67	0.06139	0
50	104	0.06899	0

A fixed maintenance schedule of a fleet of ten aircraft, each equipped with two engines, serves as the starting point. The total duration of this schedule is assumed to be ten years (3650 days). $\bar{c}$ is determined independently for the three described maintenance strategies (Reactive, Preventive, and Predictive), and for the Reference Case to provide a theoret-

ical lower limit. The five Tests Sets, TS1 to TS5, are again considered individually.

A previously defined framework (dePater et al., 2022) that focuses on the comparison of perfect and imperfect RUL prediction during maintenance scheduling shall serve as the basis. This framework is modified and extended to incorporate Preventive and Reactive Maintenance when comparing costs per cycle. Also, the total conducted maintenance tasks, the average interval between two consecutive maintenance tasks, and the total number of engine failures of each strategy are compared.

7.1. Methodology and Assumptions

The proposed framework and integer linear program for scheduling each maintenance task can be formally described as follows. Ten aircraft are considered within the framework:

$$\text{Aircraft } k \in K;\ \text{with } K = \{1, 2, 3, \ldots, 10\}.$$

Each aircraft $k \in K$ is equipped with two engines of the same type:

$$\text{Engine } l \in L_k;\ \text{with } L_k = \{1, 2\}.$$

For each aircraft, a time horizon of ten years (3650 days) is considered:

$$\text{Day } d \in D;\ \text{with } D = \{1, 2, 3, \ldots, 3650\}.$$

For simplicity reasons, it is assumed that each day d corresponds to one engine cycle. Each aircraft is scheduled to undergo an A-check within a specified interval. The A-check slots $s \in S_k$ for aircraft k are defined well in advance. According to the "Engine Maintenance Concepts for Financiers" (dePater et al., 2022; Shannon & Ackert, 2011), A-check slots are usually determined to be within an interval of $10 \ldots 20$ days. Those A-check slots for the corresponding aircraft k are defined as follows:

$$\text{Slots } s \in S_k;\ \text{with } S_k = \left\{s_{k1}, s_{k2}, \ldots, s_{ki}, \ldots, s_{k(n-1)}, s_{kn}\right\};$$

$$\text{with } 10 \leq \left(s_{ki} - s_{(k(i-1))}\right) \leq 20; \forall i \in n \text{ and } \forall k \in K$$

n describes the total number of A-checks for each aircraft. Due to the random intervals between different A-check slots, n cannot be specified and varies between different aircraft. For example, A-checks are scheduled on the following days for aircraft 1, as depicted by the green slots in Figure 21:

$$S_1 = \{11, 27, 41, \ldots, 121, 139, \ldots, 209, 220, 233, \ldots\}$$

During each A-check, the aircraft is in the hangar, and basic maintenance tasks are conducted as defined by the aircraft manual. Those tasks primarily focus on visually inspecting the aircraft structure, such as flaps, slats, control surfaces, and breaks. The engines may also be visually inspected for external damage or leakage, but any in-depth engine inspection or repair is usually not part of a regular A-check (Department for Business Innovation and Skills, 2016). This framework does not consider other, less frequent letter checks, such as B-, C-, or D-checks. The days on which an A-check for aircraft $k \in K$ occurs are determined in a way that they do not overlap with the A-checks of other aircraft due to capacity reasons:

$$S_k \cap S_{(k+1)} = 0; \forall k \in K$$

During each A-check slot $s \in S_k$, additional engine maintenance tasks for one of the engines $l \in L_k$ of aircraft $k \in K$ can be scheduled. Since the aircraft is in the hangar during this time, selecting an A-check slot to perform additional engine maintenance is particularly advantageous to reduce overall maintenance costs and maximize aircraft availability. To determine during which A-check slot s the engine l of aircraft k shall be maintained, an Integer Linear Program (ILP) is proposed. The decision variable x_{kls} is defined as follows:

$$x_{kls} = \begin{cases} 1; \text{ if engine } l \text{ of aircraft } k \text{ is maintained during slot } s \\ 0; \text{ else} \end{cases}$$

The target day d^{target} shall specify the cost-optimal cycle to conduct the additional engine maintenance task. Since the A-check schedule is defined well in advance, d^{target} may not coincide with an available A-check slot $s \in S_k$, especially for Predictive Maintenance. Therefore, a penalty score p_{kls} is proposed as the objective function. For each day the selected slot $s \in S_k$ is earlier than d^{target}, a penalty of $p^{\text{early}} = 1$ is incurred. On the other hand, for each day the selected slot $s \in S_k$ is later than d^{target}, a penalty of $p^{\text{late}} = 10$ is incurred. The resulting objective function can be described as follows:

$$p_{kls} = p^{\text{late}}\left(s - d^{\text{target}}\right) + p^{\text{early}}\left(d^{\text{target}} - s\right) \tag{17}$$

The penalty score p_{kls} must be minimized to determine the ideal A-check slot to conduct the additional engine maintenance, which results in the following objective:

$$\text{Min} \sum_{s \in S_k} p_{kls} \cdot x_{kls} \tag{18}$$

To determine d^{target}, it must be differentiated between the individual maintenance strategies:

1. **Reference Case:** As it was described in the previous chapter, it is assumed that the RUL of each engine is known without error. Therefore, d^{target} is always set to the last cycle before engine failure. Furthermore, the A-check schedule is adapted in a way that d^{target} always coincides with an available A-check slot. Again, this case shall provide a lower cost per cycle limit, which can only be determined in this hypothetical case and cannot be replicated in a real scenario due to inevitable uncertainty.

2. **Reactive Maintenance:** No engine maintenance is conducted before failure, which means the engine always fails. d^{target} is thus not determined.
3. **Preventive Maintenance:** For Preventive Maintenance, as described in the previous chapter, an optimal engine maintenance interval t^* based on the failure distribution of the individual Test Set (TS1 to TS5) is determined. Based on this optimal maintenance interval, each cycle during which engine maintenance must be conducted is considered as the target day d^{target}. Since d^{target} can be anticipated in advance for Preventive Maintenance, it is assumed that this can be incorporated into the A-check schedule. Therefore, an A-check slot always coincides with d^{target}, resulting in a minimal penalty score of $p_{kls} = 0$.
4. **Predictive Maintenance:** The RUL is continuously updated during each cycle by the proposed MLP model described in Chapter 5. As soon as the predicted RUL falls below the determined threshold T^*, the operational buffer $\beta = 20$ days is necessary to prepare the required maintenance task, i.e., order spare parts and ensure that the required personnel capacity is available. This results in the target day $d^{\text{target}} = T^* - \beta$.
As opposed to Preventive Maintenance, d^{target} is determined after the A-checks are scheduled for each engine. Since re-scheduling an A-check slot on short notice is generally undesirable due to additional costs, the slots do not necessarily coincide with the target day d^{target}. The ideal slot to conduct the engine maintenance is therefore determined by minimizing the penalty score p_{kls} according to eq. 18. Therefore, d^{target} is not a strict deadline, which would generally result in an additional, unscheduled maintenance slot with higher costs, but rather a guideline. If no A-check slot coincides with d^{target}, scheduling the Maintenance earlier than d^{target} is preferred to decrease the likelihood of an unexpected failure.

A further capacity constraint is assumed, limiting the additional maintenance tasks to **one engine per A-check slot**. This constraint can formally be described as:

$$\sum_{k \in K} \sum_{l \in L_k} x_{kls} \leq 1; \forall s \in S_k$$

Also, **only one maintenance slot shall be scheduled** for each engine simultaneously. After the additional maintenance task is conducted, a new slot can be scheduled for the following task.

$$\sum_{s \in S_k} x_{kls} = 1; \forall l \in L_k, \forall k \in K$$

The 20 engines of the considered test set (TS1 to TS5) are assigned randomly (with replacement) to each aircraft.

Each engine is assumed to begin in the "as good as new" state. When an additional engine maintenance task is conducted, the engine deterioration shall be reset to the state "as good as new", and a new engine is randomly selected from the respective TS. The behavior of this new engine (specifically, how it degrades and its ultimate failure cycle) is then assigned to the corresponding engine of the aircraft. In a real scenario, the total engine life and degradation behavior are characterized by uncertainty. The variability employed within this framework can be increased by imposing a different engine behavior on each aircraft engine after a maintenance task. This can be visualized in Figure 22: The first engine of aircraft 1 primarily mimics the behavior of engine 12 from TS1. After the first maintenance task is conducted at cycle 217 on that engine, its behavior is "replaced" by engine 15 of the same TS.

A future extension of this framework could help to determine the influence of resetting the engine to a varying state between "as bad as old" and "as good as new". Also, taking other forms of letter checks (B, C and D) into account may be a consideration.

7.2. Results and Discussion for the Maintenance Framework

The described framework shall be applied to the maintenance strategies in the following sections.

7.2.1. Reference Case

The Reference Case is considered first to determine the lower baseline for $\bar{c}_{\text{ref}}$. Figure 22 depicts the two engines of Aircraft 1 as an example: the theoretical engine 1 failure is at cycle 217 (represented by the red line). Therefore, cycle 217 is chosen to perform the engine maintenance task. Table 14 summarizes the results for all five TS. In addition to $\bar{c}_{ref}$, the number of conducted maintenance tasks, the failure count (Failures), and the average interval between two consecutive engine maintenance tasks (Avg. Int.) are provided.

7.2.2. Reactive Maintenance

On the other hand, as a "worst-case scenario", Reactive Maintenance is considered. In Figure 23, Cycle 217 is marked orange for engine 1 of aircraft 1, meaning failure occurs during this cycle. The results in Table 15 show that the average costs per cycle increase by 874.0% compared to the Reference Case. The average interval between two maintenance tasks increases slightly compared to the Reference Case.

7.2.3. Preventive Maintenance

As determined in Chapter 6, the optimal interval varies between $t^* = 119.9$ cycles for TS1 and $t^* = 125.2$ cycles for TS5, with an average of $t^* = 122.9$ cycles.

As stated above, the A-check schedule is adapted to the optimal maintenance interval of the engines, meaning an available slot s coincides with d^{target} of the engine. There is only one exception: at the beginning of this schedule at cycle 1, both aircraft engines are in the state "as good as new" and should ideally undergo the first engine maintenance task after $t^* = 119$ days for TS1. Due to the assumed capacity constraint, only one of both engines can be maintained during this A-check slot. The engine maintenance of the second engine is thus scheduled for the A-check slot with the minimal

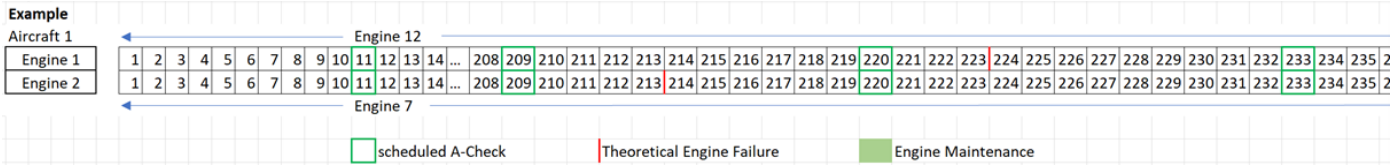


Figure 21: Graphical representation of an exemplary schedule for aircraft 1

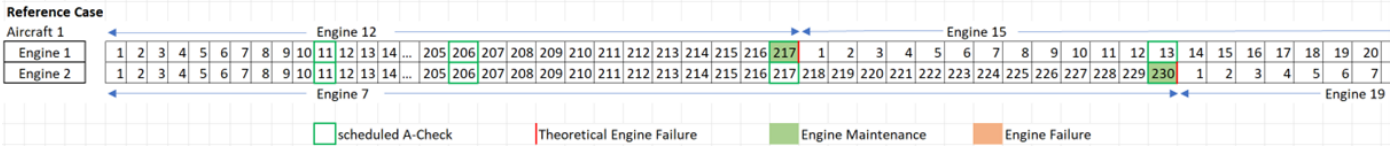


Figure 22: Graphical representation of the schedule for the Reference Case, (Aircraft 1, Engines 1 and 2)

Table 14: Maintenance Framework results for the Reference Case

Test Set	Total Costs	$\bar{c}_{\text{ref}}$	Maintenance Tasks	Failures	Avg. Int.
TS1	3410	0.04832	341	0	206.9
TS2	3880	0.05466	388	0	183.0
TS3	3470	0.04906	347	0	203.8
TS4	3490	0.04934	349	0	202.7
TS5	3170	0.04509	317	0	221.8
Average	3484	0.04929	348.4	0	203.6

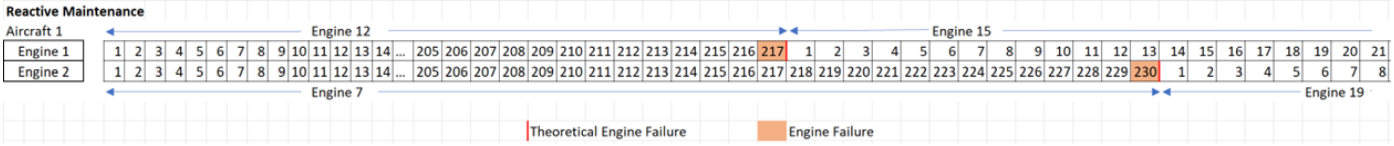


Figure 23: Graphical representation of the schedule for Reactive Maintenance, (Aircraft 1, Engines 1 and 2)

Table 15: Maintenance Framework results for Reactive Maintenance

Test Set	Total Costs	$\bar{c}_R$	Maintenance Tasks	Failures	Avg. Int.	$\bar{c}_R$ Incr. to Ref. [%]
TS1	35200	0.48527	0	344	206.1	852.6
TS2	38900	0.55598	0	394	179.9	882.6
TS3	34400	0.49038	0	344	203.9	874.7
TS4	33700	0.48633	0	342	205.6	885.7
TS5	30600	0.44551	0	315	224.5	873.6
Average	34560	0.49269	0	347.8	204,0	874.0

penalty score. This can be graphically represented in Figure 24: The engine maintenance task should optimally be performed at cycle 119. Only one engine, in this case, engine 1, is maintained during the A-check slot at this cycle; Maintenance for engine 2 of the same aircraft is scheduled for the A-check slot at cycle 109, which corresponds to the minimal penalty score p_{kls}.

For the subsequent engine maintenance slots, this overlap will not occur. The results are provided in Table 16. The average interval is slightly shorter than the optimal interval for each Test Set due to the capacity constraint. The average cost/cycle increase is 83.3% compared to the Reference Case.

7.2.4. Predictive Maintenance

The yellow marks in Figure 25 indicate the cycle at which the RUL falls below T^*, e.g., cycle 156 for engine 1. Ideally, an A-check slot in precisely 20 days would be chosen for aircraft 1 engine 1, corresponding to engine cycle 176. Since no A-check is scheduled for this cycle, the other A-checks are evaluated based on p_{kls} according to eq. 18. The following three A-checks lots after cycle 156 are scheduled for cycle 162 ($p_{kls} = 14 \cdot 1 = 14$), cycle 175 ($p_{kls} = 1 \cdot 1 = 1$), and cycle 192 ($p_{kls} = 16 \cdot 10 = 160$). Also, no additional engine maintenance for the second engine is scheduled for either of the three A-check slots yet. Therefore, the A-check

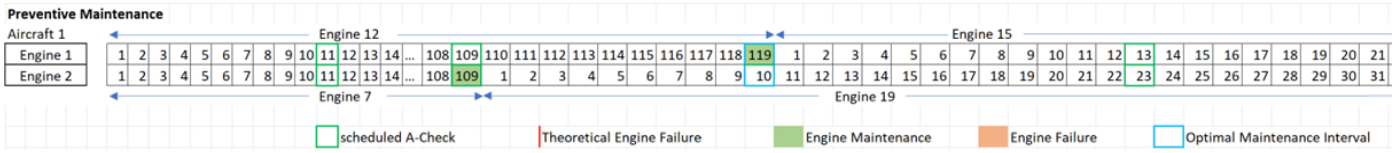


Figure 24: Graphical representation of the schedule for Preventive Maintenance, (Aircraft 1, Engines 1 and 2)

Table 16: Maintenance Framework results for Preventive Maintenance

Test Set	t^*	Total Costs	$\bar{c}_{Pv}$	Maintenance Tasks	Failures	Avg. Int.	$\bar{c}_{Pv}$ Incr. to Ref. [%]
TS1	119.9	6250	0.08973	625	0	116.7	85.7
TS2	124.7	6270	0.08961	627	0	122.3	63.9
TS3	120.5	6240	0.09016	624	0	118.1	83.8
TS4	124.3	6260	0.09050	626	0	121.8	83.4
TS5	125.2	6240	0.09169	624	0	122.7	103.4
Average	122.9	6252	0.09034	625.2	0	120.3	83.3

slot with the minimal penalty score $p_{kls} = 1$ at cycle 175 (marked green) is chosen to conduct the additional engine maintenance. The results are summarized in Table 17: $\bar{c}_{Pd}$ is on average 17.3% higher compared to $\bar{c}_{\text{ref}}$, but on average approximately 36% lower compared to $\bar{c}_{Pv}$. Also, the average interval between maintenance tasks is on average 54.4 cycles longer than for Preventive Maintenance, corresponding to an increase of around 45%.

7.2.5. Discussion

Summarizing the findings, this maintenance framework for a fleet of aircraft verifies the results of the cost comparison methodology proposed in Chapter 6. When comparing all maintenance strategies to the Reference Case in terms of $\bar{c}$, Predictive Maintenance outperforms both other strategies significantly with an average cost increase of approx. 17.3% (874.0% for Reactive Maintenance and 83.3% for Preventive Maintenance). As discussed in Section 6.2.4, reducing the operational buffer can decrease the costs for Predictive Maintenance further. Although $\bar{c}_{Pd}$ is significantly lower than $\bar{c}_{Pv}$, safety has not been compromised as no failures occur for both strategies.

In many industries, Preventive Maintenance is the most prevalent maintenance strategy. By employing Predictive Maintenance, $\bar{c}$ can be reduced by approximately 36.0%. Also, the average interval between two consecutive maintenance tasks can be increased by approximately 45%, thus maximizing engine availability.

As a further consideration, just as for Preventive Maintenance where the A-check schedule is adapted to the optimal maintenance interval t^*, similar flexibility could be allowed for Predictive Maintenance, possibly decreasing $\bar{c}_{Pd}$ even further.

8. Conclusion

Maintenance costs are a significant contributor to operating expenses. Also, a substantial percentage of maintenance costs are wasted due to unnecessary over-maintenance or improperly conducted maintenance. This can be particularly detrimental in today's globalized world, because operational processes must be optimized to be competitive.

This thesis aims to develop a methodology to determine and compare the average costs per cycle $\bar{c}$ for Reactive, Preventive, and Predictive Maintenance, using turbofan jet engine data of the NASA CMAPSS dataset as an example. Compared to previous research, this thesis combines several aspects which are usually only considered individually: first, the most suitable Machine Learning algorithm for this application is determined, and a hyperparameter tuning is conducted. Next, the methodology to compare different maintenance strategies is developed, and the average costs per cycle are compared. Lastly, a framework to apply the methodology to a realistic maintenance schedule of a fleet of ten aircraft is discussed.

The results suggest that Predictive Maintenance significantly outperforms Reactive and Preventive Maintenance in terms of $\bar{c}$. When the cost comparison methodology of Chapter 6 is considered, $\bar{c}_{Pd}$ for Predictive Maintenance is on average 30.1% lower compared to $\bar{c}_{Pv}$ for Preventive Maintenance, which is currently the most common strategy. Furthermore, when considering the described maintenance framework for a fleet of aircraft as described in Chapter 7, the costs per cycle for Predictive Maintenance are 36.0% lower compared to Preventive Maintenance and even 88.3% lower compared to Reactive Maintenance.

In 2019, global MRO spending for aircraft accumulated to $91 billion, and by assuming a 36.0% cost decrease, up to $32 billion could eventually be saved if Preventive Maintenance is replaced by Predictive Maintenance.

In addition to a decrease in MRO expenditure, employing Predictive Maintenance and monitoring the engine degradation during operation can reduce the risk of unexpected engine failure and thus increase safety. Also, the interval between successive maintenance tasks can be increased signif-

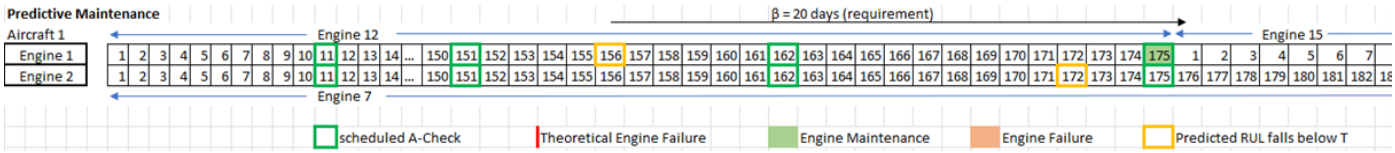


Figure 25: Graphical representation of the schedule for Predictive Maintenance, (Aircraft 1, Engines 1 and 2)

Table 17: Maintenance Framework results for Predictive Maintenance

Test Set	T^*	Total Costs	$\bar{c}_{Pd}$	Maintenance Tasks	Failures	Avg. Int.	$\bar{c}_{Pd}$ Incr. to Ref. [%]
TS1	41	3900	0.05556	390	0	184.2	15.0
TS2	54	4650	0.06400	453	0	156.2	17.1
TS3	51	3980	0.05631	398	0	176.7	14.8
TS4	64	4190	0.05990	419	0	166.8	21.4
TS5	57	3740	0.05334	374	0	188.2	18.3
Average	53.4	4092	0.05782	406.8	0	174.4	17.3

icantly, resulting in higher engine or machine availability.

It must be noted that Predictive Maintenance is the most complex and costly strategy to implement. A prior analysis if this strategy is suitable for the respective application is always required. Also, different simplifications are applied throughout this thesis, such as only considering A-checks and keeping c_f and c_m constant, regardless of the remaining engine cycles. The proposed methodology can be extended in future research to also incorporate other aspects, such as: resetting engines to varying degrees between "as good as new" and "as bad as old", varying operational buffers and cost assumptions, and different Machine Learning algorithms. Also, applying this framework to other applications can verify that the results are independent of the application.

This thesis highlights the necessary steps to determine the cost-optimal maintenance strategy. Although jet engine data is considered, this shall only serve as an example that can be applied to other applications in various industries.

References

An, A., Kim, N., & Choi, J. (2015). Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. *Reliability Engineering and System Safety*, *133*, 223–236.

Beauchamp, J. (2020). A visual comparison between the complexity of decision trees and random forests. Retrieved March 15, 2023, from https://commons.wikimedia.org/wiki/File:Decision_Tree_vs._Random_Forest.png

Carvalho, T., Soares, F., Vita, R., Francisco, R., Basto, J., & Alcala, S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering*, *137*.

Cline, B., Niculescu, R. S., Huffman, D., & Deckel, B. (2017). Predictive maintenance applications for machine learning. *Annual Reliability and Maintainability Symposium (RAMS)*, 1–7.

Dawotola, A., Trafolis, T., Mustaffa, Z., & Gelder, P. (2013). Risk-based maintenance of a cross-country petroleum pipeline system. *Journal of Pipeline Systems Engineering and Practice*, *4*.

Deighton, M. G. (2016). *Facility Integrity Management*. Gulf Professional Publishing.

Department for Business Innovation and Skills. (2016). UK Aerospace Maintenance, Repair, Overhaul and Logistics Industry Analysis. Retrieved March 15, 2023, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/502588/bis-16-132-uk-mrol-analysis.pdf

dePater, I., Reijns, A., & Mitici, M. (2022). Alarm-based predictive maintenance scheduling for aircraft engines with imperfect remaining useful life prognostics. *Reliability Engineering and System Safety*, *221*.

Diamanti, K., & Soutis, C. (2010). Structural health monitoring techniques for aircraft composite structures. *Progress in Aerospace Science*, *46*(8), 342–352.

Dodge, Y. (2008). *The Concise Encyclopedia of Statistics*. Springer.

Gilabert, E., Fernandez, S., Arnaiz, A., & Konde, E. (2017). Simulation of predictive maintenance strategies for cost-effectiveness analysis. *Proceedings of the Institution of Mechanical Engineers*, *231*(13).

Haroun, A. (2015). Maintenance cost estimation: application of activity-based costing as a fair estimate method. *Journal of Quality in Maintenance Engineering*, *21*(3).

Hu, Y., Miao, X., Si, Y., Pan, E., & Zio, E. (2022). Prognostics and health management: A review from the perspectives of design, development and decision. *Reliability Engineering and System Safety*, *217*.

Huellermeier, E. (2021). Aleatoric and epistemic uncertainty in machine learning. Retrieved March 23, 2023, from https://www.gdsd.statistik.uni-muenchen.de/2021/gdsd_huellermeier.pdf

IATA. (2021). Airline maintenance cost executive commentary, fy2021 data. Retrieved March 14, 2023, from https://www.iata.org/contentassets/bf8ca67c8bcd4358b3d004b0d6d0916f/fy2021-mctg-report_public.pdf

IBM. (2023). K-nearest neighbors algorithm. Retrieved March 15, 2023, from https://www.ibm.com/topics/knn#:~:text=Next%20steps-,K%2DNearest%20Neighbors%20Algorithm,of%20an%20individual%20data%20point

Ignatovich, S., Menou, A., Karuskevich, M., & Maruschak, P. (2013). Fatigue damage and sensor development for aircraft structural health monitoring. *Theoretical and Applied Fracture Mechanics*, *65*, 23–27.

Ihn, J., & Chang, F. (2014). Detection and monitoring of hidden fatigue crack growth using a built-in piezoelectric sensor/actuator network: I. diagnostics. *Smart Materials and Structures*, *13*(3).

Junqiang, L., Malan, Z., Hongfu, Z., & Jiwei, X. (2014). Remaining useful life prognostics for aeroengine based on superstatistics and information fusion. *Chinese Journal of Aeronautics*, *27*(5), 1086–1096.

Lee, H., & Scott, D. (2009). Overview of maintenance strategy, acceptable maintenance standard and resources from a building maintenance operation perspective. *Journal of Building Appraisal*, *4*, 269–278.

Lei, Y., Naipeng, L., Guo, L., Li, N., Yang, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical Systems and Signal Processing*, *104*.

Li, H., Zhao, W., & Zio, E. (2020). Remaining useful life prediction using multi-scale deep convolutional neural network. *Applied Soft Computing*, *89*.

Li, X., Ding, Q., & Sun, J. (2018). Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliability Engineering and System Safety*, *172*, 1–11.

McCaffrey, J. (2013). The neural network input-process-output mechanism. Retrieved March 15, 2023, from https://visualstudiomagazine.com/articles/2013/05/01/neural-network-feed-forward.aspx

Mobley, R. (2002). *An Introduction to Predictive Maintenance*. Butterworth Heinemann.

Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2017). *Machine Learning*. Tailor; Francis Group, LLC.

NASA. (2023). Cmapss jet engine simulated data. Retrieved March 15, 2023, from https://data.nasa.gov/Aerospace/CMAPSS-Jet-Engine-Simulated-Data/ff5v-kuh6

National Institute of Standards and Technology. (2022). Engineering statistics handbook. Retrieved March 23, 2023, from https://www.itl.nist.gov/div898/handbook/eda/section3/eda35.htm

Nguyen, K., & Medjaher, K. (2019). A new dynamic predictive maintenance framework using deep learning for failure prognostics. *Reliability Engineering and System Safety*, *188*, 251–262.

Reid, M. (2023). Reliability - a python library for reliability engineering. Retrieved March 23, 2023, from https://reliability.readthedocs.io/en/latest/

ReliaSoft Corporation. (2009). Preventive maintenance and the cost per unit time equation. Retrieved March 23, 2023, from https://www.weibull.com/hotwire/issue96/relbasics96.htm

scikit-learn developers. (2023). 3.2. tuning the hyper-parameters of an estimator. Retrieved March 15, 2023, from https://scikit-learn.org/stable/modules/grid_search.html

Shannon, M., & Ackert, P. (2010). Basics of aircraft maintenance programs for financiers. Retrieved March 15, 2023, from http://aircraftmonitor.com/uploads/1/5/9/9/15993320/basics_of_aircraft_maintenance_programs_for_financiers___v1.pdf

Shannon, M., & Ackert, P. (2011). Engine maintenance concepts for financiers. Retrieved March 15, 2023, from http://www.aircraftmonitor.com/uploads/1/5/9/9/15993320/engine_mx_concepts_for_financiers___v2.pdf

Silvestrin, L., Hoogendoorn, M., & Koole, G. (2019). A comparative study of state-of-the-art machine learning algorithms for predictive maintenance. *Symposium Series on Computational Intelligence*, 760–767.

Singh, J., Azamfar, M., Li, F., & Lee, J. (2020). A systematic review of machine learning algorithms for prognostics and health management of rolling element bearings: fundamentals, concepts and applications. *Measurement Science and Technology*, *32*(1).

Sirvio, K. M. (2015). Intelligent systems in maintenance planning and management. In *Intelligent Systems Reference Library* (pp. 221–145, Vol. 87).

Smola, A. J., & Schoelkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*, 199–222.

Tiddens, W., Braaksma, J., & Tinga, T. (2020). Exploring predictive maintenance applications in industry. *Journal of Quality in Maintenance Engineering*, *28*(1).

Trevisan, V. (2022). *Comparing robustness of mae, mse and rmse*. Retrieved March 23, 2023, from https://towardsdatascience.com/comparing-robustness-of-mae-mse-and-rmse-6d69da870828

US Department of Defense. (2004). Mil-std-188. Retrieved March 15, 2023, from http://everyspec.com/MIL-STD/MIL-STD-0100-0299/MIL-STD-188-141D_55865/

US Department of Defense. (2011). Mil-std-3034. Retrieved March 15, 2023, from https://web.archive.org/web/20110927135601/https://assist.daps.dla.mil/quicksearch/basic_profile.cfm?ident_number=277649

Vachtsevanos, G., Lewis, F., Roemer, M., Hess, A., & Wu, B. (2006). *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley.

Vollert, A., & Theissler, A. (2021). Challenges of machine learning-based RUL prognosis: A review on NASA's C-MAPSS data set. *International Conference on Emerging Technology and Factory Automation*, *26*, 1–8.

Wang, Y., Gogu, C., Binaud, N., Bes, C., Haftka, R., & Kim, N. (2017). A cost driven predictive maintenance policy for structural airframe maintenance. *Chinese Journal of Aeronautics*, *30*(3), 1242–1257.

Zhao, X., Gao, H., Zhang, G., Ayhan, B., Yan, F., Kwan, C., & Rose, J. (2007). Active health monitoring of an aircraft wing with embedded piezoelectric sensor/actuator network: I. defect detection, localization and growth monitoring. *Smart Materials and Structures*, *16*(4).

Junior Management Science 9(3) (2024) 1836-1858

Junior Management Science

www.jums.academy
ISSN: 2942-1861

The Munich Entrepreneurial Ecosystem in the Health Sector: Current State and Improvement Areas

Alexandra Hanna James

Technical University of Munich

Abstract

The study explores entrepreneurial ecosystems, which emphasize the impact of regional context on entrepreneurship, with a particular focus on the health sector. Given its innovation and knowledge intensity, coupled with industry-specific challenges, the health sector serves as an ideal case for ecosystem analysis. This research aims to qualitatively assess Munich's health sector entrepreneurial ecosystem and to identify actionable recommendations for enhancement. To this end, 15 interviews with entrepreneurs, venture capitalists, angel investors and support organization members are analyzed through computer-assisted qualitative content analysis. Results reveal strengths in demand, talent, knowledge, and intermediaries, while physical infrastructure, ecosystem leadership, and formal institutions score lower. The study provides concrete improvement ideas in the areas of financial support, incubators and networks, entrepreneurial education, availability of information and industry collaboration. These insights can be used to strengthen and expand Munich's entrepreneurial ecosystem, contributing to economic, societal, and technological advancements.

Keywords: biotechnology; digital health; entrepreneurial ecosystems; entrepreneurship; life science

1. Introduction

In recent years, entrepreneurship research has shifted from focusing specifically on the individual entrepreneur to also considering the wider context of entrepreneurship - the entrepreneurial ecosystem.[1,2] This approach centers around the influence the specific regional context exerts on the entrepreneurial process, providing a systems perspective on entrepreneurship.[3] More specifically, entrepreneurial ecosystems can be defined as "a set of interdependent actors and factors coordinated in such a way that they enable productive entrepreneurship within a particular territory".[4]

Efficient entrepreneurial ecosystems should be of major importance to policymakers, scholars, and practitioners, as an efficient ecosystem facilitates higher entrepreneurial activity, which in turn results in economic growth and job creation.[5] The impact of entrepreneurial ecosystems can be summarized in three categories: economic, technological, and societal. Economic impacts relate to the increase in regional wealth, prosperity, and reputation. The technological impact refers to the regional technological innovation that is achieved by the actors of the ecosystems, including new ventures, universities, and research institutions. Societal impact can be described as non-monetary outcomes that are beneficial for society, such as the creation of new products and services.[6]

One industry which has been highly relevant in the past decades is biotechnology and the greater health sector. The National Venture Capital Organization (NVCA) even consid-

I would like to thank my supervisor Dr. Svenja Jarchow for her invaluable guidance, support and insightful feedback throughout the entire process of completing this master thesis. A special acknowledgement also goes to the interview partners who generously contributed their time and expertise to this project.

1 See Stam (2015, pp. 1759-1763).
2 See Spigel and Harrison (2018, p. 151).
3 See Stam and van de Ven (2021, p. 809).
4 Stam and Spigel (2018, p. 407).
5 See Stam and van de Ven (2021, p. 810).
6 See Audretsch et al. (2019, pp. 317-319).

DOI: https://doi.org/10.5282/jums/v9i3pp1836-1858

ered biotech to be the second most important industry both in terms of investments as well as quantity of deals, the information technology (IT) sector being number one.[7] The amount of global biotech venture funding has risen significantly during the past decade and reached $23 billion in 2020, as shown in Figure 1.[8] The digital health industry also displays rapid growth, with a projected compound annual growth rate (CAGR) of nearly 27% until 2030.[9] Since the health sector is highly innovative and knowledge is frequently renewed, networks are becoming increasingly important to distribute new knowledge and best practices.[10] This dependence on networks among companies and entrepreneurs in the health sector makes it an interesting industry to study entrepreneurial ecosystems.

Germany has a highly innovative pharmaceutical, biotech, medical technology, and IT sector. Being home to over 400 companies in medtech and over 270 companies active in digital health, the sectors generated € 11.4 billion and € 400 million in revenue in 2018, respectively.[12,13] The state of Bavaria employs the second highest number of biotech professionals, second only to North Rhine-Westphalia. Munich alone has two biotechnology clusters, the BioM Munich Biotech Cluster and Cluster Biotechnology.[14] Additionally, there are five Digital Health Hubs and Accelerators, including the German Accelerator Life Science/IT and the Digital Health Accelerator.[15] Since the foundation of the BioM in 1997, Munich has developed into a major biotech and health technology center. With two leading universities, the Technical University of Munich (TUM) and the Ludwig Maximilian University of Munich (LMU), and several research institutes such Max-Planck-Institute for Biochemistry, the city also possesses the necessary institutions to provide talent and knowledge.[16]

Entrepreneurial ecosystem research has thus far remained mostly industry agnostic. It is often argued that the benefits of an ecosystem are mostly related to entrepreneurship-specific knowledge, rather than industry-specific knowledge and is therefore relevant to a broad range of industries.[17] However, in the health industry, start-ups face significant barriers to success due to the nature of the industry, namely very long development processes, a strict regulatory environment and complex technologies.[18] Therefore, entrepreneurs in the health sector, including biotech, medtech and digital health, can significantly benefit from both entrepreneurial and industry-specific knowledge in an ecosystem. Since the metropolitan area of a city is generally viewed to be the most appropriate level of analysis for entrepreneurial ecosystems, the health industry in Munich was chosen as research focus.[19]

The objective of this study is to qualitatively assess the entrepreneurial ecosystem in the health sector in Munich. The central research question is "How can the Munich entrepreneurial ecosystem in the health sector become more efficient?". For this purpose, 15 experts were interviewed, who work within the entrepreneurial process, either as venture capitalists or business angels, support organizations or as entrepreneurs themselves. Based on this analysis, the aim is to develop an understanding of the quality of the different elements of the Munich ecosystem and discover actionable improvements and recommendations to advance the ecosystem.

In the first chapter, I briefly review the literature on the main concepts of entrepreneurial ecosystems, which underpin this study. In the second chapter, the main characteristics of the health industry are explained in more detail. To put the opinions of key stakeholders into perspective, it is vital to understand the distinguishing features of this industry. I then outline the methodological approach used to conduct and analyze the interviews, before describing the results. Next, I discuss the implications of the results, the limitations of this study and key recommendations for ecosystem improvements. The final chapter concludes and presents avenues for future research.

2. Main concepts of entrepreneurial ecosystems

In the following, I introduce the main concepts of entrepreneurial ecosystems, including the elements of an ecosystem and the most frequently used framework, governance approaches, the life cycle theory of ecosystems and critique of the concept.

2.1. Elements of entrepreneurial ecosystems

Entrepreneurial ecosystems combine all elements necessary to facilitate entrepreneurship in a particular region. There have been several proposed frameworks to conceptualize the aspects of entrepreneurial ecosystems. An early approach by Spigel (2017) summarized entrepreneurial ecosystems in three types of attributes: cultural, social, and material. This approach is depicted in Figure 2. According to this concept, cultural attributes refer to a supportive culture and histories of entrepreneurship. Building on this, social attributes such as networks, talent, mentorship, and investment capital facilitate the resources and means to create new ventures. Finally, material attributes encompass policies, universities, infrastructure, open markets, and support services. These attributes should not be seen as distinct layers, but overlapping factors, which support and reinforce one another.[20]

[7] See NVCA (2021, pp. 28-29).
[8] See Senior (2021, p. 408).
[9] See Grand View Research (2022, pp. 1-2).
[10] See Lechner and Dowling (1999, p. 320).
[11] Taken from Senior (2021, p. 408).
[12] See Germany Trade & Invest (2018a, pp. 2-9).
[13] See Germany Trade & Invest (2018b, pp. 2-9).
[14] See Germany Trade & Invest (2018b, pp. 2-9).
[15] See Germany Trade & Invest (2018a, pp. 2-9).
[16] See Lechner and Dowling (1999, p. 321).
[17] See Spigel and Harrison (2018, p. 162).
[18] See Baeyens et al. (2006, p. 31).
[19] See Leendertse et al. (2021, p. 478).
[20] See Spigel (2017, pp. 50-57).
[21] Modified, taken from Spigel (2017, p. 57).

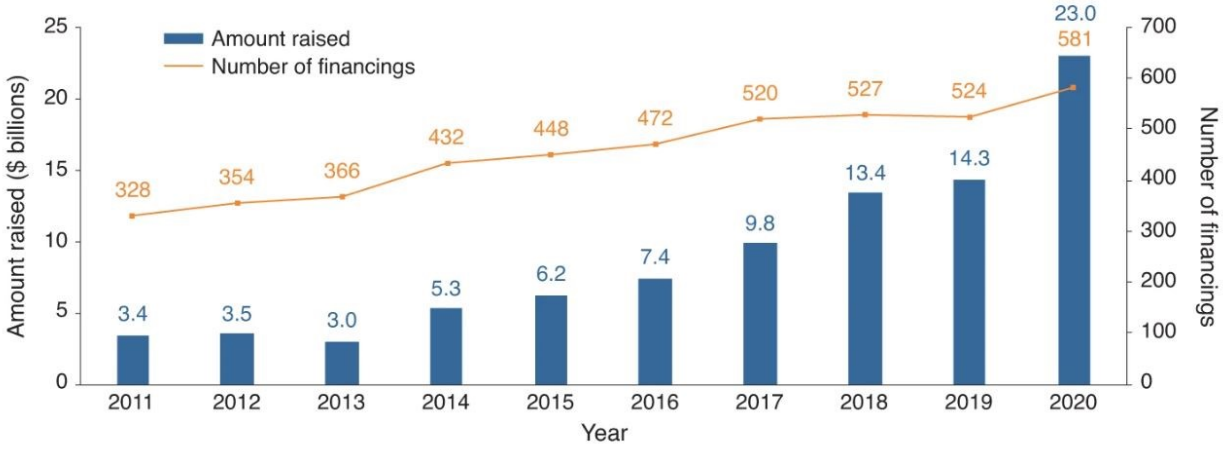


Figure 1: Global venture capital investments in the biotech sector 2011-2020[11]

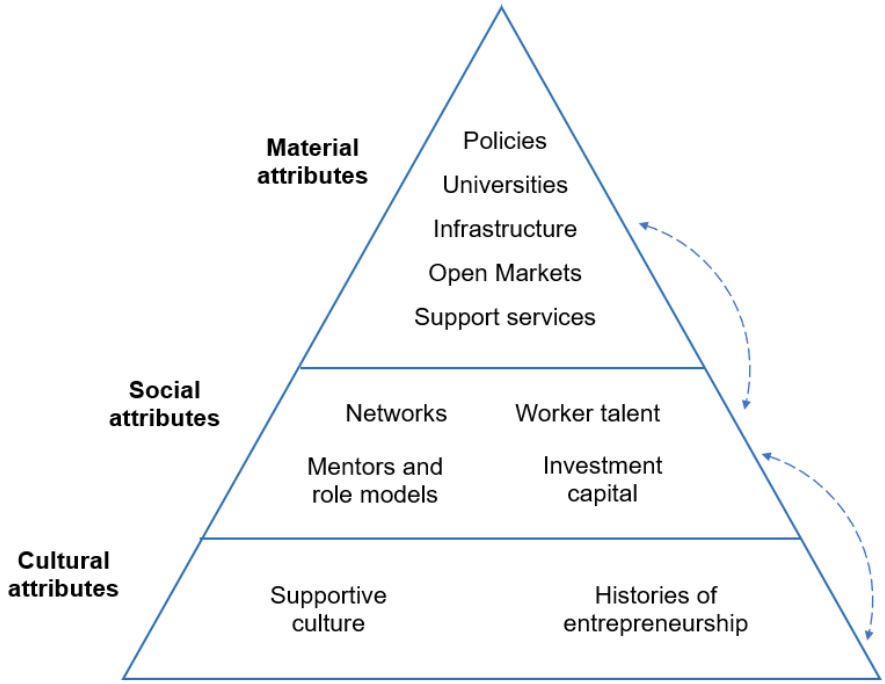


Figure 2: Relationships between the attributes of entrepreneurial ecosystems[21]

This framework was later modified and expanded by Stam and van de Ven (2021), who summarized entrepreneurial ecosystems in an integrative model including ten elements, which is depicted in Figure 3.

In this framework, the ten elements which together form the ecosystem can be divided in two distinct layers: institutional arrangements and resource endowments. Institutional arrangements are regarded to be the formal institutions, culture, and networks, which underpin the ecosystem. Formal institutions, such as the regulatory framework, represent the fundamental precondition for entrepreneurship. Formal institutions therefore guide and regulate economic action, the acquisition and use of resources and the entrepreneurship process. This element also includes educational institutions, the healthcare system, and law enforcement. The culture element encapsulates the attitude towards and perception of entrepreneurship in society. A supportive entrepreneurial culture can be described as one which highly values entrepreneurship and normalizes the risks and challenges associated with the entrepreneurship process. This not only encourages entrepreneurs to create new ventures, but also increases the willingness of skilled individuals to

[22] Modified, taken from Stam and van de Ven (2021, p. 813).

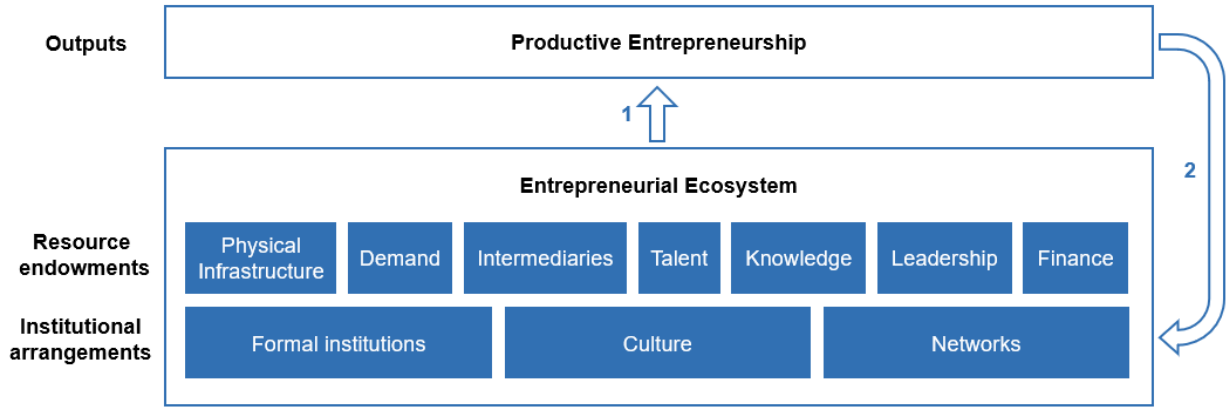


Figure 3: An integrative model of the elements and outputs of entrepreneurial ecosystems[22]

work for a young start-up. As a final element in the institutional arrangements layer, entrepreneurial networks facilitate collaboration between the different stakeholders and the distribution of knowledge, talent, and capital within the ecosystem.[23,24]

Taking a closer look at the second level, the resource endowments refer to physical infrastructure, demand, intermediaries, talent, knowledge, leadership, and finance. The element of physical infrastructure describes the accessibility of the region in terms of transportation, but also the availability of physical space, including office space and laboratories, and the digital infrastructure. Demand refers to the general purchasing power and market size for the novel products and services, which are developed by the entrepreneurs.[25] Intermediaries or support services can help young ventures with specific challenges and thereby lower the entry barriers for entrepreneurs, increasing the speed of innovation. Examples of intermediaries include incubators, accelerators, or other support services such as legal advice or consultancies. The element of talent refers to the availability of skilled individuals, both in terms of education and previous experience in the entrepreneurial space or in the industry. Knowledge from private and public organizations and the investment in the development of new knowledge is crucial for the process of innovation and therefore a further vital element in the framework. Leadership provides guidance in the ecosystem by the presence of interest groups and experienced leaders, who can mentor younger entrepreneurs. The final element in the resource endowments layer is finance. The accessibility and supply of finance for long-term and uncertain entrepreneurial projects is a vital resource for successful entrepreneurship.[26,27]

In combination, these ten elements facilitate productive entrepreneurship and the value it creates as output, which is portrayed by arrow one in the ecosystem framework. A proxy that is often used to measure productive entrepreneurship is the prevalence of high growth firms in an area. Stam and van de Ven (2021) found a very strong positive correlation between the strength of the ten elements and the quantity of high growth firms in a particular region and thus concluded that the overall quality of the entrepreneurial ecosystem is positively related to entrepreneurial output. Furthermore, their results showed that the ten elements are highly interdependent and co-evolve within a region. Therefore, the elements should be viewed as a whole system to explain the levels of entrepreneurial activity in a specific region.[28]

Finally, entrepreneurial output feeds back into the ecosystem, which is represented by arrow two in the framework. The research group identified a positive correlation between the prevalence of high-growth firms and subsequent values of the individual elements of the ecosystem. This positive feedback can be explained by two main reasons. First, successful entrepreneurs often share their experience with the ecosystem by becoming e.g., venture capitalists, mentors, or network developers.[29] Second, the existence of entrepreneurial activity encourages the creation of new ventures by signaling that entrepreneurship is a legitimate and potentially a successful pathway.[30]

This framework relies on the assumption that all elements are of equal importance in entrepreneurial ecosystems. Several researchers argue that this is a rather agnostic approach as it can be reasoned that certain elements are more important than others in shaping an ecosystem.[31] Corrente et al. (2019) investigated the weighting of elements and proposed that some factors are more critical for the success of entrepreneurial ecosystems than others. The research team did not use the same ten elements shown above, however, their analysis suggested that “Culture and Social Norms” is the

[23] See Spigel and Harrison (2018, p. 155).
[24] See Stam and van de Ven (2021, pp. 813-815).
[25] See Leendertse et al. (2021, p. 482).
[26] See Stam and Spigel (2018, p. 415).
[27] See Stam and van de Ven (2021, pp. 813-817).
[28] See Stam and van de Ven (2021, pp. 827-828).
[29] See Stam and van de Ven (2021, p. 827).
[30] See Kuckertz (2019, p. 477).
[31] See Leendertse et al. (2021, p. 483).

most important factor, followed by "Government Programs" and "Internal Market Dynamics".[32]

2.2. Governance of entrepreneurial ecosystems

The governance of entrepreneurial ecosystem is often described as a challenging topic because ecosystems are highly complex structures. In addition, the interdependency between the various elements of an ecosystem makes it difficult for public initiatives to target specific areas for improvement in a particular ecosystem. In addition, governance initiatives to promote entrepreneurship have been found to frequently fail, because strategies that have been successful in other regions are simply replicated in seemingly similar locations.[33]

To guide the governance of entrepreneurial ecosystems, researchers have identified several characteristics, which must be considered when devising initiatives to encourage entrepreneurial output: entrepreneurial ecosystems are self-regulating, highly complex and location specific. Firstly, ecosystems are dynamic, self-regulating networks and behave in an unpredictable manner. Governing ecosystems, i.e., changing the self-regulating mechanisms, is therefore extremely challenging, as the outcomes of interventions are difficult to foresee.[34,35] Secondly, ecosystems consist of various factors and stakeholders, interacting in a complex manner. Entrepreneurs are frequently considered to be the main drivers of the ecosystem; they are however only one factor among many. As previously described, all elements are highly interdependent and co-evolve, therefore if one element is weak, it forms a bottleneck, and the development of the entire ecosystem can start to stagnate. Only if these weaknesses can be overcome, the ecosystem can function effectively and promote entrepreneurial activity.[36] Finally, ecosystems are highly location specific and shaped around the local circumstances, such as networks, culture, or geographic location. An approach, which works for one particular ecosystem therefore cannot simply be copied and applied to another location.[37]

Building on these ecosystem characteristics, governance principles suggested by current entrepreneurial ecosystem literature can be summarized into three main concepts. First, governance initiatives should support self-regulation of the ecosystem in a bottom-up approach rather than governing the ecosystem in a top-down approach. This is grounded in the fact that entrepreneurial ecosystems develop mostly through the complex and uncoordinated interactions of the individual actors.[38] Within this stakeholder network, entrepreneurs should act as the central leaders of the ecosystem, as they are best equipped to recognize its possibilities and limitations. In addition, radical interventions in entrepreneurial ecosystems can have unforeseen consequences, due to their dynamic nature. Therefore, any governance initiative must acknowledge and cultivate the inherent evolutionary potential and support the self-regulation of the ecosystem in a minimally invasive way. As ecosystems behave and react in an unforeseeable manner, it is also recommended to act incrementally to avoid over-steering the ecosystem.[39] Stam (2015) summarizes the role of ecosystem governance "as a feeder of the ecosystem rather than as a leader".[40]

Secondly, creating impactful and effective policies for entrepreneurial ecosystems requires a holistic approach that considers its complexity. As outlined in the entrepreneurial ecosystems framework, these systems consist of different elements, which are all connected and interdependent. Focusing on all stakeholders and their connections therefore reduces silo thinking and improves the overall quality of the ecosystem.[41] Lastly, it is important to preferentially focus on weaknesses of the entrepreneurial ecosystem, rather than trying to further improve the strengths. As mentioned, ecosystems are characterized by bottlenecks, which must be overcome for the system to be efficient at promoting entrepreneurship. The holistic perspective implies that all factors are important in creating an environment conducive to entrepreneurial activities. Therefore, when allocating resources to an ecosystem, its weakest elements should be targeted first to alleviate any bottleneck situations.[42,43]

Emphasizing these concepts, Isenberg (2010) published nine principles, which should be followed when governing an entrepreneurial ecosystem. In the first five principles, he focuses on the role of local characteristics and the bottom-up process of governance: 1. Stop emulating Silicon Valley, 2. Shape the ecosystem around local conditions, 3. Engage the private sector from the start, 4. Stress the roots of new ventures and 5. Do not over-engineer clusters; help them grow organically. According to these principles, it is crucial to focus on and support already existing local structures, industries, and trends in contrast to creating entirely novel ones. In addition, the author stresses the importance of the long-term and profit-driven motivation and perspective of the private sector in developing self-sustaining ecosystems. In the following three principles, he focuses on creating a culture, which supports ambitious entrepreneurship: 6. Favor the high potentials, 7. Get a big win on the board and 8. Tackle cultural change head-on. Isenberg suggests favoring the most ambitious and growth-oriented ventures to not only maximize the creation of wealth, but also inspire potential future entrepreneurs and reduce the perception of entrepreneurial risks and barriers. With the last principle, he stresses the importance of institutions: 9. Reform legal, bureaucratic, and

[32] See Corrente et al. (2019, p. 513).
[33] See Colombelli et al. (2019, p. 505).
[34] See Kuckertz (2019, p. 478).
[35] See Colombo et al. (2019, pp. 427-428).
[36] See Audretsch et al. (2016, pp. 373-375).
[37] See Audretsch et al. (2016, p. 360).
[38] See Roundy et al. (2018, pp. 8-10).
[39] See Kuckertz (2019, pp. 479-480).
[40] Stam (2015, p. 1761).
[41] See Kuckertz (2019, p. 479).
[42] See Audretsch et al. (2016, p. 373).
[43] See Kuckertz (2019, p. 479).

regulatory frameworks. [44]

Ultimately, ecosystem policy and initiatives can promote entrepreneurial activity, but only under the premise that the basic conditions in the structure of the ecosystem are met. By providing resources to ecosystems, they can be strengthened, but without efficient networks between entrepreneurs, these resources were shown to have only a limited effect.[45]

2.3. Life cycle of entrepreneurial ecosystems

As entrepreneurial ecosystems emerge and develop, they change significantly and typically enter various phases. The lifecycle of ecosystems can be summarized in five stages: the emergence of an ecosystem, the growth phase, the stabilization or maturity phase, the decline phase and finally, a re-emergence phase. The birth phase of an ecosystem starts with an idea leading to new venture creation. This early entrepreneurship can either result from an employee or academic exploiting knowledge overlooked by the incumbent firm or by spin offs when incumbent firms or universities encourage individuals to seize opportunities by creating a new company. This first phase is characterized by a low number of company birth rates in the area and the formation of entrepreneurial networks between them. The subsequent growth phase is characterized by the increased pace of new entrants. In addition, intermediaries start to offer entrepreneurship specific programs and financial capital becomes more easily available. While talent becomes more entrepreneurially minded, this is also the most important bottleneck and source to grow. The end of the second phase is marked by the tendency of incumbent firms to re-integrate startups.[46,47]

In the stabilization phase firm birth rates are declining, and an increasing number of firms are acquired by incumbents. Other ventures might mature and become more structured and less dynamic. In addition, market opportunities, networks and the entrepreneurial culture starts to weaken. In this phase, effective leadership in the ecosystem is crucial to sustain its development. The decline phase is characterized by a low rate of new venture creation. New ideas and technologies are mainly explored within established firms. However, this is not necessarily accompanied by a decline in overall wealth or competitiveness of the region. The ecosystem leadership potentially reorients itself towards other economic development initiatives. Finally, the region might enter a re-emergence phase, in which the life cycle starts again in an accelerated way. Since the intermediaries, entrepreneurial culture and networks are already in place, entrepreneurs can more easily start new companies and commercialize new ideas.[48]

With each lifecycle stage of an entrepreneurial ecosystem, its ideal governance design also changes. In the birth stage, the first entrepreneurial ventures often act as catalysts for the creation of an ecosystem. These central actors often initially govern the processes and collaboration in the new ecosystem. Private institutions and support networks rotate around the central actors, providing crucial resources for the sustainment of the ecosystem. Therefore, early ecosystem governance can often be described as rather hierarchical. Once the ecosystem grows and expands, governance typically shifts to a more horizontal governance design, where multiple stakeholders interact and collaborate without the direction of a central player. As entrepreneurial networks become increasingly dense and more actors take central positions within the ecosystem, governance is based on implicit understandings, unofficial routines, and shared norms. Finally, the governance consists of a well-connected network of actors, which together shape and develop the ecosystem.[49]

2.4. Critique of the entrepreneurial ecosystems concept

Despite its recent popularity, the concept of entrepreneurial ecosystems was critiqued for three main reasons: being tautological, showing no clear cause and effect and a questionable level of analysis. First, the concept was claimed to be rather tautological: entrepreneurial ecosystems support the creation of new ventures and wherever there are a high number of successful ventures, there is evidently a good entrepreneurial ecosystem. This relationship links back to the positive feedback successful entrepreneurs have on the local ecosystem. It has been argued that such tautological reasoning ultimately provides little insights to inform public policy or further research.[50] Second, the concepts of entrepreneurial ecosystems so far merely provide long lists of relevant elements without a clear explanation of cause and effect. There is still no universal consensus about the definition of the ecosystem itself and the coherence and causal interdependent effects of the individual framework elements. While the elements do provide some focus, a more consistent explanation of the framework would provide clearer guidelines for further research and ecosystem governance. To improve this ecosystem explanation, it is necessary to differentiate between the essential and contingent elements of an ecosystem and more clearly define the ideal role of the regulatory institutions and other public organizations.[51] Finally, it is still unclear which general scope and level of analysis is most appropriate with regards to entrepreneurial ecosystems. The boundaries of ecosystems can be defined geographically, focusing on either a country, a region, or a city. Ecosystems could also be defined based on industries or corporations, which offer opportunities for venture creation and growth.[52]

[44] See Isenberg (2010, pp. 42-49).
[45] See Spigel and Harrison (2018, p. 162).
[46] See Cantner et al. (2021, pp. 413-417).
[47] See Mack and Mayer (2016, p. 2123).
[48] See Cantner et al. (2021, p. 417).
[49] See Colombelli et al. (2019, pp. 508-511).
[50] See Stam and Spigel (2018, p. 415).
[51] See Alvedalen and Boschma (2017, pp. 893-895).
[52] See Stam and Spigel (2018, pp. 415-416).

3. Challenges in the health sector

The health industry comprises firms in multiple sectors, such as pharma, (digital) healthcare, biotechnology, medical devices, and diagnostics. Ventures in the health sector often face significant barriers to success by the nature of the industry, namely the long development processes, regulatory difficulties, and highly complex technologies.

3.1. Technological complexity

First, the technologies and product development processes in the health industry tend to be highly complex. For the pharma, biotechnology, and medtech sectors, understanding them in detail often requires deep scientific knowledge of molecular biology and engineering techniques and processes.[53]

The digital health sector is a further field in the health space, which is home to an increasing number of start-ups including digital therapeutics, health information technology, telehealth and telemedicine, smart devices, and personalized medicine. Digital therapeutics products, for instance, deliver therapeutic interventions to patients via software solutions to prevent, manage or treat a medical disorder or disease. These digital healthcare products frequently integrate advanced technologies, such as machine learning or artificial intelligence, to optimize the treatment of patients and their subsequent health outcomes.[54]

Therefore, investors might have difficulties understanding the technology and the industry environment when evaluating ventures in this industry. Generalist venture capital (VC) firms frequently outsource the due diligence of ventures in the health industry since they lack the capabilities to adequately assess the business model internally. Even specialist investors often miss the expertise in the specific field, which makes it difficult to analyze the investment opportunity and, consequently, challenging for entrepreneurs to secure funding.[55]

3.2. Regulatory complexity

Second, entrepreneurial ventures in the health space need to comply with numerous rules and regulations across all functions of their company, as they are operating in a highly regulated industry. This includes assessments of product quality, clinical and research design, patient safety, but also navigating ethical issues around their products, financial compliance, and training of employees.[56]

Companies in the biotechnology and pharmaceutical area must navigate the regulations for market authorization of drugs in accordance with the International Code of Harmonization. After successfully completing the necessary clinical tests and trials, they must then file an application with the European Medicines Agency or the relevant German agency to receive the approval to license their product.[57] For medical devices, the introduction of the Medical Device Regulation in 2021 imposed strict requirements for post market surveillance of the launched products, next to rigorous controls and high expectations regarding the clinical data collected.[58] For digital health applications (Digitale Gesundheitsanwendung, DiGA) for the detection, monitoring, treatment or alleviation of medical conditions, Germany introduced a new approval process in 2019. To be approved, the DiGA developer must prove a positive healthcare effect by conducting a scientific comparative study.[59]

Within the entire health industry, the increasingly complex regulatory environment represents a serious challenge for young start-ups, especially in an industry where non-compliance likely has significant effects on costs, reputation, and ultimately, patient's lives.

3.3. Long development process

The health industry is characterized by a long development process until a technology is converted into a market ready product. In the pharmaceutical and biotechnology sectors, the entire process from the early discovery phase to market entry takes 15 years, on average. Given the long time to market, these companies require a very high upfront investment to cover all costs connected to the development process. Furthermore, the long development process increases the chances that superior technologies emerge on the market while the product is still under development. These factors make the industry extremely risky.[60] For medical devices, the process from concept to market takes an average of 3-7 years, where higher risk products are subject to more stringent regulatory processes compared to lower risk products.[61] The development process of digital health products and applications is more variable, although these products must also be validated by clinical studies and subsequently approved. Therefore, the time to market is still significantly higher compared to other industries.[62]

4. Methodology

In the following section, the research methodology is outlined. This includes the selection of the method, selection of interviewees and the setting of the interviews. For the subsequent analysis, the method for transcribing and analyzing the interviews is explained. Finally, I briefly discuss the employed quality criteria.

4.1. Expert Interviews

4.1.1. Selection of the method

To study the entrepreneurial ecosystem in Munich, semi-structured expert interviews were chosen as the empirical

[53] See Baeyens et al. (2006, pp. 31-35).
[54] See Dang et al. (2020, pp. 2209-2211).
[55] See Baeyens et al. (2006, pp. 32-35).
[56] See Deloitte Centre for Health Solutions (2015, pp. 3-8).
[57] See Price Water House Coopers (2009, pp. 17-18).
[58] See Maresova et al. (2015, pp. 1508-1510).
[59] See BfArM (2020, pp. 7-8).
[60] See Baeyens et al. (2006, pp. 31-33).
[61] See van Norman (2016, p. 278).
[62] See Dang et al. (2020, pp. 2210-2211).

method. Interviews are one of the most frequently used methods in the field of qualitative research, as they allow for subject-relatedness of the research as well as a detailed description and interpretation of the respective research area.[63] The flexibility of qualitative methods such as interviews makes them particularly suitable to analyze complex phenomena such as the multi-level interactions and dynamics of ecosystems.

Qualitative interviews can be distinguished according to the degree of structure. For this study, a semi-structured approach was chosen, where the first part of the interview was guided by the entrepreneurial ecosystem framework. The second part of the interview was less structured and only guided by a small selection of key questions. In contrast to an unstructured interview, better comparability between the interviewees can be achieved and it is ensured that all relevant aspects of the research question are addressed.[64]

Expert interviews are systematic and theory guided interviews with individuals who have exclusive knowledge about a particular topic.[65] Therefore, interviews were conducted with individuals who possess exclusive knowledge about the entrepreneurial ecosystem in the health area in Munich.

4.1.2. Selection of the interview partners

For the type and quality of information obtained through expert interviews, the selection of interview partners is a decisive influencing factor. Expert knowledge is traditionally tied to a function or professional role. Thus, experts are defined by their position and status as well as the knowledge which is attributed to them.[66]

To capture different perspectives in the ecosystem, entrepreneurs, investors, and support organizations were interviewed, in a similar distribution to Spigel (2017), who interviewed 70% entrepreneurs, 15% investors and 15% others, such as economic development officials.[67] Of the 15 interviews conducted for this study, 9 were with entrepreneurs (60%), 4 with venture capitalists or angel investors (27%) and 2 with support organizations of the ecosystem (13%).

Comparability between the interviews was achieved by selecting experts in similar positions in the different organizations. Interviewees in the entrepreneur category had all (co-)founded at least one Munich-based company in the industry in the last 10 years. The investors actively invest in biotech, (digital) health or medtech start-ups in Munich. For the support organizations category, the managing directors of two biotech support organizations in Munich were interviewed.

4.1.3. Setting and course of the interviews

The expert interviews were conducted in April and May 2022. The interviews were held via the web conferencing tool Zoom. The main part of the interview, excluding the introduction and concluding comments, was audio recorded by the Zoom recording function. Each interview started with a brief introduction, followed by a short explanation of the entrepreneurial ecosystem theory and the framework. Finally, the use of the data was briefly explained, and the interviewees were asked whether the interview could be recorded.

Usually, a guideline is used for semi-structured interview methods. This serves as a structuring and steering tool and represents a link between already existing theory and empiricism.[68] The interview guideline contained the entrepreneurial ecosystem framework and its ten elements. The interviewee was asked to rate the elements on a scale from 1-10 (10 being the highest score) and elaborate on the rating. This was asked to get a tentative understanding for the ecosystem as a whole and be able to better compare the quality of the different elements. In addition, the guideline contained open-ended questions to which the interviewee could respond freely at their own discretion. The order of the questions was flexible and merely served as a guideline during the interviews. At the end of the interview, there was time for a short debriefing and any questions from the interviewee. In total, the interview duration ranged between 25 and 45 minutes.

4.2. Qualitative analysis

4.2.1. Transcription

The analysis of the interviews was started by transcribing every recording. In the first step, the transcription software Happy Scribe was used. The software automatically generated a transcript from the recording, which is generally 80% accurate. To increase the accuracy of the software, approximately 30 words which were used frequently in the interviews were added to its vocabulary manually, such as “entrepreneurship”, “pharma” or “venture capital”.

In the second step, the generated transcripts were edited manually to further increase the accuracy, insert any segments which were not recognized by the software, and smoothen the text. For this purpose, duplicate words, half sentences, and filler words were deleted. Interview pauses, voice inflections and other non-verbal elements were not taken into account. Finally, the transcripts were exported with timestamps and speaker names.

4.2.2. Qualitative content analysis

The analysis of the interview content was completed using the computer-assisted qualitative data analysis software Maxqda. To analyze the content of the interviews, the methodological approach by Kuckartz was followed. Kuckartz describes three main methods of qualitative content analysis. For this study, the content-structuring analysis was chosen, which can be considered the core method of qualitative content analysis and is displayed in Figure 4. In

[63] See Mayring (2016, pp. 20-25).
[64] See King et al. (2019, pp. 52-60).
[65] See Kaiser (2014, p. 6).
[66] See Kaiser (2014, pp. 37-38).
[67] See Spigel (2017, p. 59).
[68] See Misoch (2015, pp. 65-68).

Table 1: Overview of interview partners

Interviewee	Category	Job Title	Area
E1	Support Organization	Managing Director	Biotech
E2	Investor	VC Partner	Biotech
E3	Entrepreneur	Co-founder	Digital Health
E4	Investor	VC Partner	Biotech
E5	Entrepreneur	Founder and CEO	Biotech
E6	Investor	VC Principal	Biotech
E7	Support Organization	Managing Director	Biotech
E8	Entrepreneur	Co-founder and CEO	Digital Health
E9	Entrepreneur	Co-founder and CTO	Digital Health
E10	Entrepreneur	Founder and CEO	Digital Health
E11	Entrepreneur	Founder and CEO	Medtech
E12	Entrepreneur	Co-founder and CEO	Digital Health
E13	Entrepreneur	Founder and CEO	Biotech
E14	Investor	Managing Director and Angel Investor	Healthcare, Biotech
E15	Entrepreneur	Co-founder and CEO	Medtech

this method, the material is typically coded in several phases with deductively and inductively formed categories.[69]

First, the interview transcripts were imported into Maxqda and sorted into three groups based on the type of interviewee: entrepreneurs, investors, and support organizations. After reviewing the initial results, the transcripts were reorganized into the three groups biotech, medtech and digital health, as it became apparent that the categorization along industries displayed greater differences. The first step of the Kuckartz process is initial work with the text and the creation of memos. This step was abbreviated, as the interviews had already been reviewed during the transcriptions. Next, two main categories were defined as "ecosystem status quo" and "ecosystem improvements". The ten elements of the entrepreneurial ecosystem framework were added as deductive codes to the status quo main category. In addition, a main category "other" was created, to capture any important text sequences which did not fit into any existing codes and to collect passages which indicate any limitations in the research method. The categories were then sequentially assigned to the text sections.

To develop inductive categories for the "ecosystem improvements" main category, all assigned text passages were compiled in one table. Following this, potential subcategories were collected in an unordered list before they were structured and summarized to form the final subcategories. Before the entire list of documents was coded, a trial run with 20% of the data was conducted to test the categories with regards to their applicability to the empirical material. Following this test, precise category definitions were added for every single code, consisting of a category description, the application of the category, and an anchor example showcasing a specific text passage of that category. Finally, the entire set of interviews were coded applying the inductive codes as described in the category definition.

The content analysis of the interviews was conducted with a vertical orientation focusing on the different categories. To analyze each category, the applicable text passages were filtered using the retrieved segments tool or displayed in the summary grid of Maxqda. In addition, the subcode statistic was used to display the number of interviewees who mentioned a particular subcode to gauge the importance of the respective topic. The rating of the individual ecosystem elements on a scale from 1 to 10 was summarized in a box and whisker plot. If the interviewee gave a range of numbers, the mean of this range was used for subsequent calculations. The calculation of quartiles was performed exclusive the median. Due to the relatively small sample size of 15, the results of this analysis are not generalizable or representative of the entire ecosystem in Munich. Finally, with the help of qualitative and quantitative cross tables displaying relevant text passages separately for each group, the individual groups could be compared in a systematic manner.

4.2.3. Quality criteria

The quality of empirical research is assessed using specific quality criteria. The three quality criteria, which are most often employed in qualitative research are intersubjectivity, reliability and transparency. For intersubjectivity, the subjectively obtained results are made plausible to outsiders by adequately reflecting them. The researcher's opinion must not be presented as the only correct one and offer different interpretations for readers. This criterion was achieved by clearly outlining the research rationale and reflecting the methodology and results in the discussion section. Reliability refers to the soundness of a measurement method. Since it was impossible to calculate an intercoder reliability, the retest reliability was determined. After the initial coding of the inductive subcategories, three interviews, which correspond to

[69] See Kuckartz (2020, pp. 129-132).
[70] Modified, taken from Kuckartz (2020, p. 133).

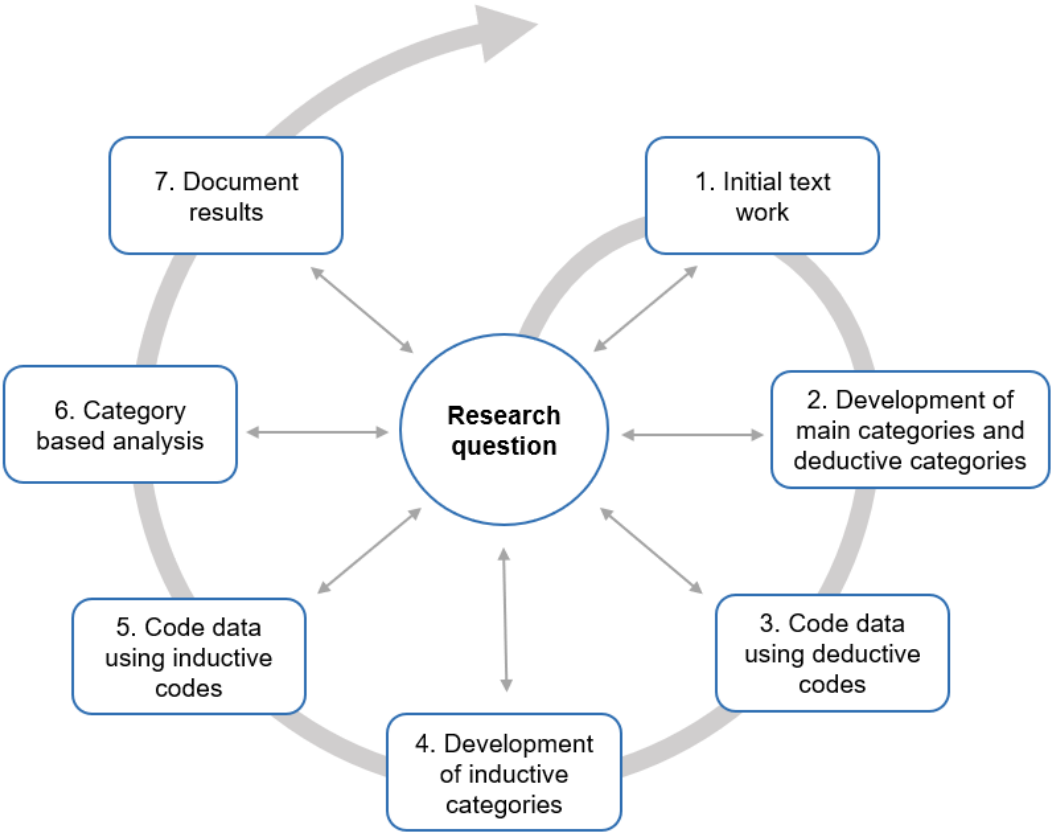


Figure 4: Process of the qualitative content analysis according to Kuckartz[70]

20% of the data, were recoded under the same conditions and the results were compared. The coded sections corresponded almost completely to the previous version except for the length of 10 coded segments. The main parts of the statements, however, perfectly matched. Transparency was achieved by documenting all important steps and thus making them comprehensible to readers. This quality criterion overlaps with the quantitative quality criterion of validity. By precisely documenting the procedure, the relevance of the method becomes apparent.[71,72]

5. Results

5.1. Ecosystem status quo

In the following section, the results relating to the current state of the Munich ecosystem are described along the ten elements of the ecosystem framework. The element scores by the interviewees are summarized in Figure 5. The elements can be categorized into three groups according to their mean score. Physical infrastructure, leadership and formal institutions were classified as low score elements, as they scored below six. Finance, culture, and networks received mean scores between six and seven and were thus grouped together as intermediate score elements. Finally, talent, intermediaries, knowledge, and demand scored 7 or above and were thus classified as high score elements.

5.1.1. Low score elements

Starting with physical infrastructure, most experts reported that affordable laboratory and office spaces are difficult to find in Munich. Seven out of the eight biotech experts think that lab space is scarce and can represent a bottleneck for founders looking to start or grow their business. Newly built lab space in incubators is oftentimes too large and therefore expensive for young companies (E5). In addition, it is challenging to secure a space: for the Innovation and Start-Up Centre Biotechnology (Innovations- und Gründerzentrum Biotechnologie, IZB) in Munich, 60 start-ups are currently on the waiting list and the waiting period is estimated to be three years (E1). Larger incubators often also prefer later stage companies to avoid rental losses (E7). E1 describes the situation as follows:

> "In my opinion, there is a huge gap here, which also prevents the teams from making progress. At the IZB in Martiensried alone, the waiting list currently has 60 start-ups that want to get in, and

[71] See Flick (2019, pp. 474-483).
[72] See Kuckartz (2020, pp. 234-237).

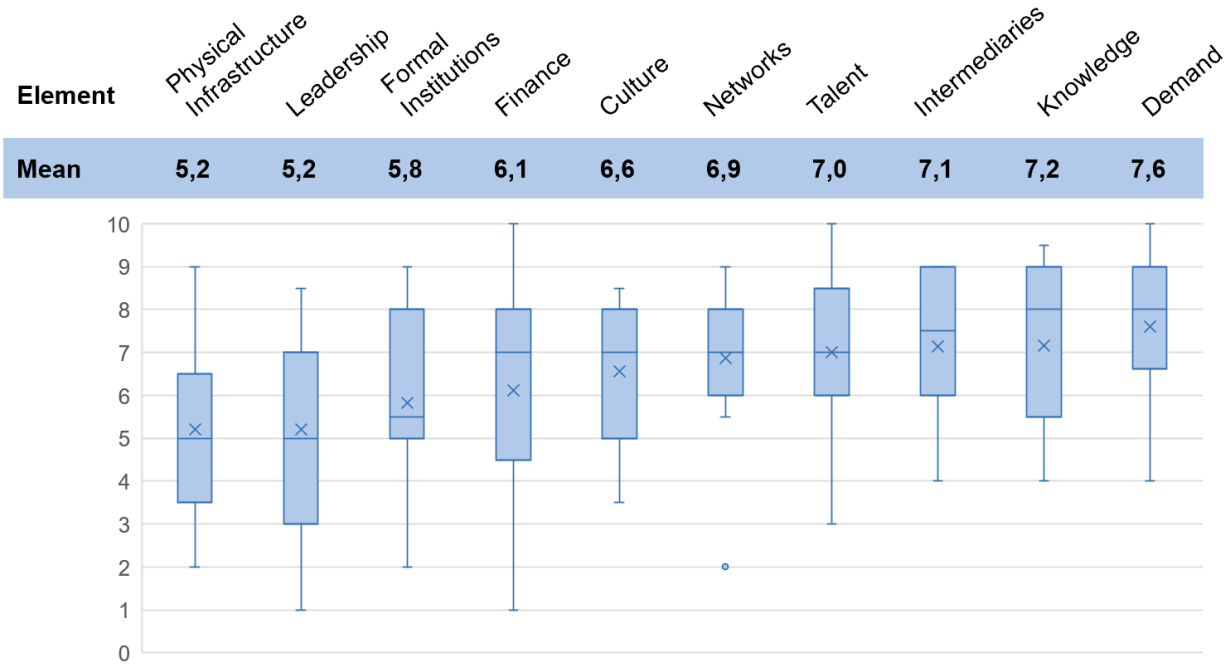


Figure 5: Box and whisker plot of the 15 interviewee's rating of the ecosystem elements

> the waiting time is said to be three years, if someone is interested now, and we all know what the time cycles are like in start-ups. In my eyes, space is a big problem in Munich. And in any case, to achieve output, something certainly needs to be done." (E1)

Novel programs like the TUM Venture Lab Healthcare are attempting to counteract this problem by offering office and lab space to founders (E6, E7). The digital health and medtech experts also mention the high cost of office space in Munich, although they oftentimes have lower requirements compared to biotech companies. The high cost of living in Munich represents an additional challenge to young entrepreneurs, which often lowers the risk one is willing to take to start a venture (E2, E10).

Leadership also represents a scarce resource in the health ecosystem in Munich. As E1 and E14 point out, the feedback loop from productive entrepreneurship back into the ecosystem does not function properly yet, as not enough companies have gone public or failed yet. For this continuous feedback, the critical mass must be greater, and in Munich the start-up wave, especially in the digital health space, is still at the beginning (E8, E14). Although there are some successful founders, they are often not available as mentors and supporters (E1, E7, E14). The experts often draw the comparison to the USA, where a significantly higher number of successful entrepreneurs share their learnings with the ecosystems and act as angel investors and mentors (E6, E8). E8 also comments that successful founders tend to move to Berlin at some point, where the ecosystem is larger compared to Munich. E15 and E14 describe leadership in Munich as follows:

> "Very bad, far too little. Some people even come to me to hear how to set up a healthcare start-up, and I'm really still at the beginning". (E15)

> "I think there are simply not many of them in total. Everything is still relatively young. There are some who are very, very successful, but they are not available every day as mentors. Of course, you have a lot of people who tried something, but have now ended up in another job because it did not work out. But a lot of them are still trying to figure out whether it will work. They are in their third or fourth or fifth year. So, finding mentors definitely needs the most improvement." (E14)

Formal institutions in Germany, such as the regulatory framework and public support programs were also often mentioned as a main pain point for entrepreneurs. Legal processes are very bureaucratic, time intensive and paperwork heavy:

> "I think Germany is very, very difficult in that respect. We have also experienced that. You are kept from starting a business, because you have to do so much next to it. You would prefer to focus on the product or the customers, however you have so much bureaucracy that you have to deal with." (E3)

> "It is essentially the case that for a small company, if you are talking about tax law, things are

> difficult to manage. So, it's simply the administrative framework, and in this respect, I would describe this part as a nuisance, as an obstacle that we are pretty much left alone with." (E5)

In addition, E2, E5 and E11 claimed that the tax system is suboptimal for young companies and systems in the Netherlands, or the UK are more accommodating to start-ups. In terms of official support programs, there are various funding and support programs in Munich and Bavaria (E1, E4, E6, E8). However, these institutions often have certain structural weaknesses, such as the absence of incentive structures which encourage innovation. In Germany, employees of such programs mostly earn a civil service salary, opposed to the USA, where these employees are entrepreneurially incentivized and receive carry as in VCs (E1). However, the Munich ecosystem seems to be more efficient when it comes to certain processes. In comparison with other German ecosystems, E4 points out that processes such as starting a company or interacting with the commercial register is significantly quicker in Munich:

> "Certain processes that are super bureaucratic are perhaps not quite so bureaucratic in Munich. Setting up a company or interacting with the commercial register, for example, is simply three times faster here than in Berlin, just as an example. So, there are certain things that are easy here, that the local network has understood." (E14)

5.1.2. Intermediate score elements

The elements finance, culture and networks were rated between six and seven, on average. With regards to finance, there is a solid venture capital industry based in Munich:

> "Munich is considered the center for biotech venture capital. That is not Berlin, that is not Hamburg, that is Munich." (E2)

E6 mentions an increase in the number of biotech funds in Germany and that many funds are starting to invest earlier. Next to VCs, Munich is home to a good angel network due to the high number of wealthy individuals (E8). In sum, most experts agree that young ventures with a solid idea and good team can in most cases attain seed financing. For later stage financing rounds, however, several experts claim that the financing landscape in Germany is suboptimal, and companies often move to different countries:

> "The biggest problem in Europe, not only in Germany, is that the stock exchanges are not here for the companies. That means that the complete cycle of capital is not closed in Germany. This means we finance companies with state money, then we add private money. And if they then want to acquire growth capital, the company needs to go to the US. This means that the technology we have financed, the know-how we have built, is transferred to the other country and that is also where the roll-out of the product takes place." (E4)

> "For the later growth phase, Series B, there are many companies that have huge problems and must resort to international investors. And because we do not have any late-stage investors, the profits are always realized abroad." (E15)

Some entrepreneurs also experienced problems connecting with investors and finding the right investor for the right project, especially if they were not part of entrepreneurial networks (E9, E14). Especially biotech founders also claimed that the risky nature of the business can make it challenging to secure even seed financing (E5, E9, E12). Finally, there are several options of public funding, such as EXIST or Bayern Kapital, which, however, have also been described as slow and relatively risk averse (E5, E15).

The entrepreneurship culture depends largely on the defined scope. Most interviewees agree on this and confirm, that the entrepreneurial culture is very strong in the various networks and programs in Munich:

> "Culture is such a bubble. I would say that within our bubble the culture is really good. But I wouldn't say that it has already arrived in society at large." (E1)

> "But I would say that the culture in which we founded the start-up i.e., in the CDTM space, is very, very good. There is a lot of support and openness for new things." (E3)

Outside of this community, society is still often skeptical of entrepreneurial projects and prefers professions with higher safety, especially in scientific disciplines. This is also reported for academic scientific and medical programs, where students are discouraged from starting their own business. E2 and E6 describe this culture as follows:

> "One of the first compliments I got when I started at Bayer was when people asked me where I worked. They said "Oh, nice, safe job". So that's what people think of first, a secure job. And of course, you can't build an entrepreneurial ecosystem with that." (E2)

> "My professors used to give me a disparaging look at university when I said I wanted to start a company. That is something which is not really accepted here yet." (E6)

However, several experts mentioned a paradigm shift, leading to more openness towards entrepreneurship in society (E6, E7, E9, E15). Due to prominent examples in the sector, more people are aware of the fact that entrepreneurship plays an important role in our economy:

> "We are sensitized to the fact that start-ups can become an important pillar of society. [...] And I hope, for example, that we have seen from the example of BioNTech that when we have a company that really takes off, it really does have a society-changing effect. [...] I think we have a great ecosystem with entrepreneurship to raise awareness that something like this is possible." (E15)

In terms of networks, there are a couple of very strong and supportive entrepreneurial networks in Munich with many meetups and conferences, however, they do not reach all entrepreneurs. There is also a clear divide between digital start-ups and healthcare and life science start-ups. For the digital sector, there are many networks, such as the Center for Digital Technology and Management (CDTM), TechFounders and Plug and Play. All five digital health founders agree that Munich has extremely powerful and supportive networks (E3, E8, E9, E10, E12); as E8 states:

> "When I came to Munich twelve years ago, I knew that I wanted to start a company someday and connecting with like-minded people was surprisingly quick in Munich if you are keen." (E8)

On the other hand, the biotech and medtech founders feel that the networks in this sector are not strong enough. Next to the BioM and the newly founded Venture Lab Healthcare there are not many offers:

> "There is something, but it can definitely be stronger to get to the point that we would say okay, there is productive entrepreneurship here." (E1)

Finally, entrepreneurs in all areas think that the networks do not reach everyone, and the options can be diffuse, fragmented, and small scale, which can make it confusing for entrepreneurs to find the right support system (E5, E10, E14).

5.1.3. High score elements

The elements talent, intermediaries, knowledge, and demand were all rated above 7, on average. The common consensus on talent among the experts was that due to highly ranked universities with yearly increasing student enrolment and prestigious research institutes, there are many well-educated scientists and management students (E1, E5, E6). Especially at the TUM and LMU, there are individual professors and chairs which specifically promote entrepreneurship (E6). The only exception is the area of computer science, where talent is a very scarce resource (E9). Due to the high presence of high-tech companies and research institutes, talent also often stays in Munich after university. There is also a mindset shift in that skilled individuals are more willing to work for a young start-up, despite it being a riskier option compared to conglomerates (E6, E8). However, E15 mentioned that it can still be very difficult to recruit skilled talent, since many professionals prefer the working conditions in larger companies such as a higher salary and fewer hours and value the freedom and responsibility in a start-up less.

Intermediaries were also rated very positively by most interviewees. E1 described the intermediary landscape as follows:

> "Support structures, especially in the early phase, be it for accounting, for financing, but also for freelance consultants, CROs and CDMOs, are becoming more established." (E1)

Within certain ecosystem structures, such as the CDTM, intermediaries were described as very strong with a lot of support available (E3). There were only two negative aspects mentioned. First, due to the plethora of intermediaries it can be difficult to find the most appropriate services for a specific stage (E9, E12). And second, many of the intermediaries, such as accelerators, are operated by non-entrepreneurs (E5, E15). E15 described his experience as follows:

> "They're not in your shoes, they can't understand it, they're not really interested, and we were also unfortunate with the mentors we have there." (E15)

The scientific, technological and entrepreneurship knowledge was praised by most interviewees. As a result of the universities LMU and TUM, the various research institutes in the city, such as the Max Planck Institute and the Helmholtz Center, and the abundance of innovative companies, there is a lot of knowledge tied to the respective groups and chairs:

> "We actually have a lot of knowledge. Above all, it feels like new chairs and research groups are formed every semester on precisely this topic, so bioinformatics with life science. We have an excellent basis with LMU and TUM as elite universities here plus all the collaborations that we have at the biotech locations like Martinsried." (E9)

Lastly, demand was the element with the highest average score. Most experts stated that Munich has a very extensive industry in the health sector and therefore has many potential cooperation partners and customers:

> "We also have an extremely good industry in Munich. That means we have the target cooperation partners right on our doorstep" (E6)

> "The demand in Munich is higher compared to other regions or large cities in Germany. Many companies are simply a step ahead. They are a bit more modern, and often simply larger with the many DAX companies that are based here." (E10)

The only critique expressed by several experts was that companies are not always willing to cooperate with start-ups. The healthcare sector, for instance, is overburdened already, and therefore healthcare professionals often do not have time to test new solutions. Other teams are overwhelmed by the many projects being implemented, especially regarding digitization, and are therefore skeptical when it comes to new digital products (E12, E14, E15).

5.2. Ecosystem improvements

Next to assessing the ecosystem, the interviewees were asked whether they could formulate any improvement ideas for the Munich ecosystem. The answers were summarized into five main areas: financial support, incubators and networks, entrepreneurial education, availability of information and industry collaboration. The number of experts who had suggestions in each area is displayed in Figure 6.

5.2.1. Financial support

The most frequently mentioned improvement area was financial support. The suggestions in this area were summarized in three main categories, which are shown in Figure 7: early-stage financing, entrepreneurship incentives and public funds. The numbers in the subcode statistic refer to the number of experts who mentioned a specific topic.

In total, nine experts advocated the improvement of early-stage financing in Germany. According to the entrepreneurs, traditional support programs are very bureaucratic, and it can take up to 18 months until the financial means are received. This means that projects are significantly slowed down or cannot be started. It was often mentioned that entrepreneurs should have easier access to smaller grants to support the initial ideation phase:

> "I could imagine that it would be helpful if you could get little money with less effort to be able to do proof of concept." (E5)

E2 suggested diversifying the standard entrepreneurship support programs, depending on the individual idea. The standard twelve to eighteen months are simply too short for many projects in the life sciences. Therefore, according to E2, it would be more appropriate to have a spectrum of funding periods for different projects. E7 suggested implementing a solution like the Small Business Innovation Research programs in the US. With such a program, the company could first apply for a small amount of funding and subsequently must prove how they utilize the money. In a second round, they are eligible to receive higher amounts of funding, which is also when business angels often join. E8 mentions that while EXIST is a great program, it is very academically driven. In his view, it would be appropriate to create programs which reward smaller funds to enable an initial ideation phase or proof of concept, even in a non-academic context. E12 suggests the creation of a separate public fund for health start-ups provided by the Ministry of Health. It is argued by E12 that the yearly healthcare costs in Germany amount to € 400 billion, so setting up a fund of € 500 thousand to promote start-ups in the health sector, which could lead to innovation and cost efficiencies, would be a reasonable option. Finally, several experts think that the ecosystem needs more support for young professionals or students to enable them to develop ideas and start businesses during their studies:

> "For individuals who have a sound idea, contribute € 800/900 every month for two or three years. That's € 10.000 a year, that's not a lot of money. But for a student, that's the difference between trying something out and waiting tables." (E14)

E14 calls this type of proposed support program "Entrepreneurial Bafög" and points out that it could be tied to a mentorship program. In return for this funding, recipients would be obligated to participate in a buddy program when they have completed their studies. This could entail sharing their experience and lessons learned with the ecosystem, even if they have not become successful founders. This program could also be established in different formats, depending on the size of the program and the amount of funding. With regards to larger programs, such as EXIST, this could include contributing as a sparring partner in other projects.

In addition to early-stage financing options, four experts mentioned ideas which involve creating financial incentives to start a business, especially in the early stage, when it is too early for private investors. As joining a start-up involves a lot of risk, some experts suggest counterbalancing this with tax incentives:

> "As an employee of a young company you are exposed to an increased risk because the company might no longer exist in two years. This is not rewarded in terms of taxation." (E4)

E2 and E4 both mention simplification of the tax on stock options for young companies. In addition, E10 suggests support programs in terms of discounted office spaces or equipment. E7, on the other hand, suggests altering the amortization policy for the invested capital:

> "If I make a loss in this sector, I can't write it off against profits I have made in the real estate sector, for example. [...] This is regulated differently in other countries. Yes, that is perhaps where I still see the greatest need." (E7)

Finally, two experts discussed possibilities to improve the structure and functioning of public grants. Especially, they mentioned that public support programs could be much more closely linked to the private sector, including VCs and larger companies:

> "I would like to see the formal institutions, if they include funding programs by the government or the states, to be much more closely linked to the

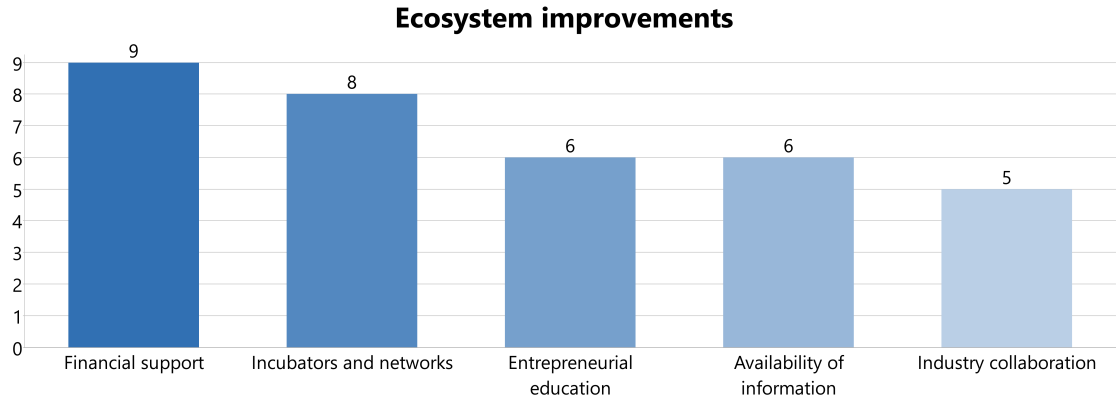


Figure 6: Subcode statistic ecosystem improvements

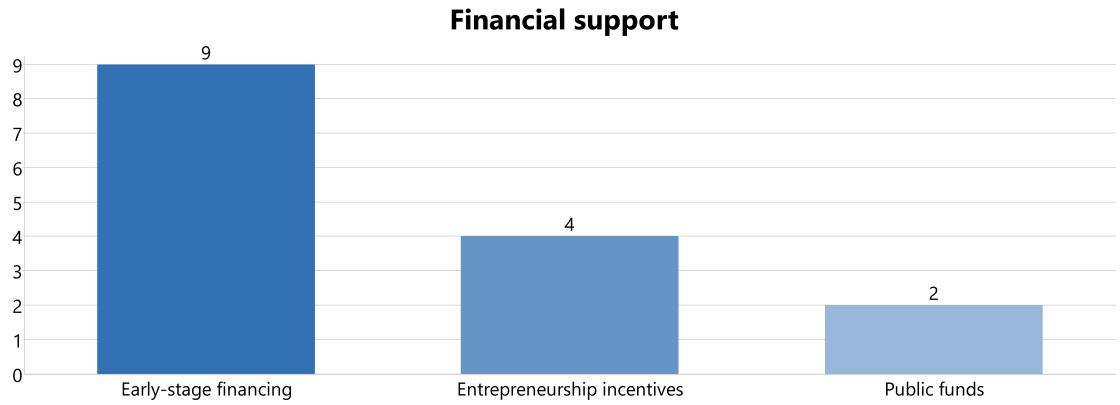


Figure 7: Subcode statistic financial support

> private sector. I mean private, whether it is venture capital or biotech companies or pharmaceutical companies, but in such a way that projects can be promoted that actually have a realistic chance, because we often think that millions are squandered where you know from day zero that nothing will be released to the market." (E2)

They suggest that a counterbalance is developed, including representatives from the private sector in the decision making. In addition, it is suggested that these committees are incentivized in a different way, similar to the incentive structure in a VC:

> "In the private sector, the company's money is your money. And then you make decisions very differently. And you can incentivize people in such a way that they make these decisions much more attentively or much more forward-looking." (E2)

5.2.2. Incubators and networks

The second improvement area was mentioned by eight experts and concerns incubators and networks. The suggestions were classified into the two areas networks and mentorship and incubator refinement, as shown in Figure 8.

Seven experts suggested improving the current networks and mentorship programs in Munich. E1 highlights the fact that due to the high complexity of the healthcare and life science sector, strong networks and the exchange of knowledge is critical. It is also important for aspiring founders to know that a support network of other founders and intermediaries exists. E6 states that these networks should include founders, patent lawyers, regulatory consultants, and experienced industry professionals to, for instance, identify the right experiments. For this network to function, E6 also adds that incentive structures need to be in place to motivate participants. In terms of implementation, it is proposed to create a campus structure where these project teams and different network participants can meet. Several experts would also prefer more coaching by founders, who experienced the process firsthand. For instance, E8 describes that programs and classes teaching fundraising and finance were often taught by VCs, who were mainly interested in getting information from the start-ups. The interviewee suggests integrating more founders, who have experienced the same journey:

> "Letting more independent founders speak at events would probably be helpful. Founders who have gone through all the phases and who could point out exactly what is available." (E8)

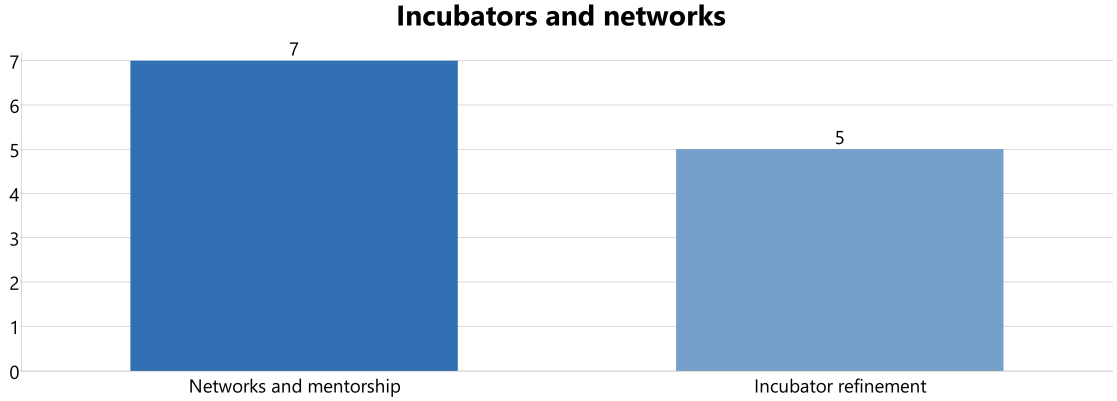


Figure 8: Subcode statistic incubators and networks

E5 agrees with this and comments that founders have a lot more relevant practical knowledge and could, for instance, teach other entrepreneurs how to connect with investors. And finally, as already mentioned in the financial support chapter, E14 suggested combining public funds with a mentorship program, where funded entrepreneurs agree to engage in a support network to share their experience with the next generation of entrepreneurs.

In addition, five experts had ideas how incubators could be improved and expanded. Since affordable lab space is scarce in Munich, it was suggested by four experts to create shared lab space, for instance in the form of an incubator (E5, E6, E7, E8). These labs should ideally be equipped with materials and devices to enable young teams to test ideas easier. Sharing this space with other teams would significantly reduce the cost and additionally lead to an increased exchange between the teams. For instance, E7 talks about the effect of collaboration during an international bootcamp:

> "I must say, through the conversations that the people and the teams have with each other, they get more know-how than through the advice of experts. So, you can learn quite well from the mistakes made by others." (E7)

Furthermore, E7 calls for a cross-sectional incubator to connect the life sciences with medtech and IT. Finally, E9 suggests expanding incubators to accommodate for more participant batches and perhaps even specialized batches with more specific topics, for instance clinic automation.

5.2.3. Entrepreneurial education

One improvement area mentioned by six of the interviewees is entrepreneurial education. This includes increased entrepreneurial education during and potentially even before university and improving entrepreneurial culture in general. The number of experts who mentioned each point is displayed in Figure 9.

Firstly, four interviewees thought that entrepreneurial education could be improved during university programs, especially for medical and scientific disciplines (E1, E6, E8, E11). E1 explains that even though students in these degree programs would be predestined to start companies in this field, this aspect is not covered at all in their studies. E1 and E6 therefore both suggest that students in scientific and medical programs could be made more aware that even basic research cannot always be altruistic but should also lead to product development. E11 also proposes teaching entrepreneurial competences across more degree programs and not merely in business studies. He further suggests interdisciplinary offers for PhD programs for students who are interested in starting a company:

> "TUM is the entrepreneurial university. You could really consider whether you could incorporate more interdisciplinary elements into the curricula." (E11)

Finally, E11 suggests starting lecture series with the start-up scene or bringing start-up fairs directly to the university campus to naturally promote the exchange between academia and founders.

The two experts E4 and E8 suppose that entrepreneurial education could be promoted even before university, starting in secondary school. E4 explains that this education should be started early on to create an understanding of entrepreneurship in school and apprenticeships. E8 agrees and states that these programs must not only be aimed at university students but should be started earlier in the various secondary schools:

> "I cannot see any reason why you have to pursue an academic career to understand that you can also found." (E8)

Finally, three interviewees mentioned the need to improve the entrepreneurship culture in the ecosystem (E4, E5). They suggest that by already starting entrepreneurial education in school, a better understanding of entrepreneurial values could be created. These include personal responsibility, risk awareness and distancing ourselves from the general culture of envy. Furthermore, failure should not be viewed

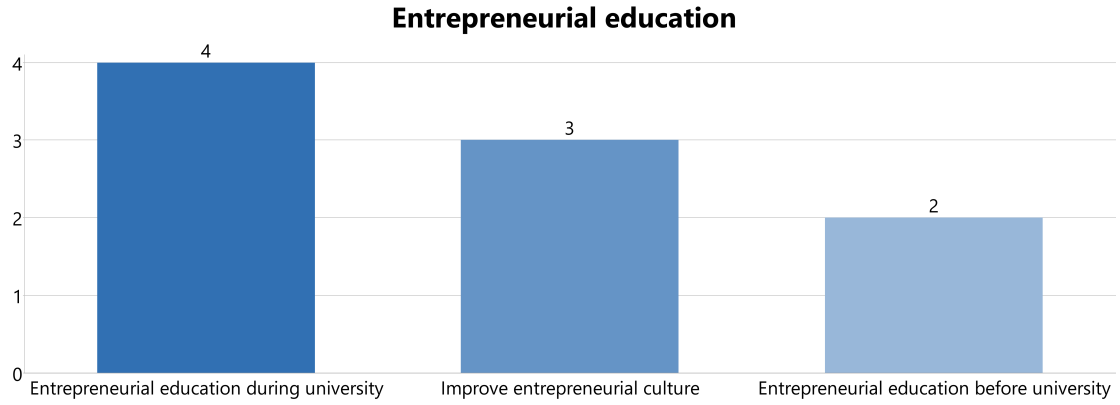


Figure 9: Subcode statistic entrepreneurial education

as a stigma, but should be viewed as a normal process and an opportunity to learn and improve. They explain these points as follows:

> "Someone who achieved something must be seen in a positive light. This starts very early and here we must improve a lot." (E4)

> "It must be completely normal for someone to totally fail with the first two companies and then it just becomes the third." (E5)

To improve the culture, E6 suggests highlighting successful ventures of the health sector even more, as they are in the tech industry:

> "You don't see it much in biotech or in health care, even though it has such a huge impact on all of our lives. That should also happen much more." (E6)

5.2.4. Availability of information

It was highlighted by six experts that the availability of information could be improved in the Munich ecosystem. This includes the creation of an information platform and an overview of the available networks, as shown in Figure 10.

Four experts suggested creating a platform with information regarding topics such as taxes, hiring and cooperation partners (E3, E5, E8, E9). E5 explains that it is extremely difficult to find contacts for certain practicalities, for instance payroll accounting and tax advisors, as a young business, especially without groups such as the CDTM or Manage&More. E9 adds, that every start-up experience bears similar problems and challenges. At the moment, this information is exchanged repeatedly between individual founders or in networks:

> "You have to click through the networks and ask and ask and ask until maybe someone knows something. Sure, it's the normal process, but as I said, many of these questions are repeated every few months by different founders. You could probably also bundle them together somewhere." (E9)

E3 and E9 suggest an open-source platform, which can be edited and expanded by founders and contains a list of contacts for certain areas and general best practices. In addition, E8 thinks the information on financing options for each respective start-up phase provided by independent founders would be extremely useful.

Furthermore, four experts propose a greater transparency with regards to which networks and support programs are available to start-ups (E3, E8, E10, E14). At the moment, the knowledge of certain networks is mostly conveyed via word of mouth or in specific entrepreneurship programs and therefore does not reach every start-up. E14 suggests linking the public support programs such as public grants more closely with the networks:

> "You could perhaps oblige every program or network supported by state or regional funds to be listed on some kind of marketplace. Then the young, committed people can educate and orient themselves." (E14)

In addition, E10 emphasizes that start-ups which are not associated with an accelerator program or university require more guidance which networks are available. He suggests that this information could be provided when founders interact with institutions anyway, for instance when registering a business.

5.2.5. Industry collaboration

Finally, five experts advocated improvements with regards to the collaboration with industry partners. The improvement points were categorized into two areas: improving the connection with industry and creating a standard framework for collaboration, as shown in Figure 11.

Several experts argue that, although Munich has an extensive health sector related industry and therefore high demand, these companies are often not accessible to start-ups. E14 suggests that existing institutions such as the Chamber

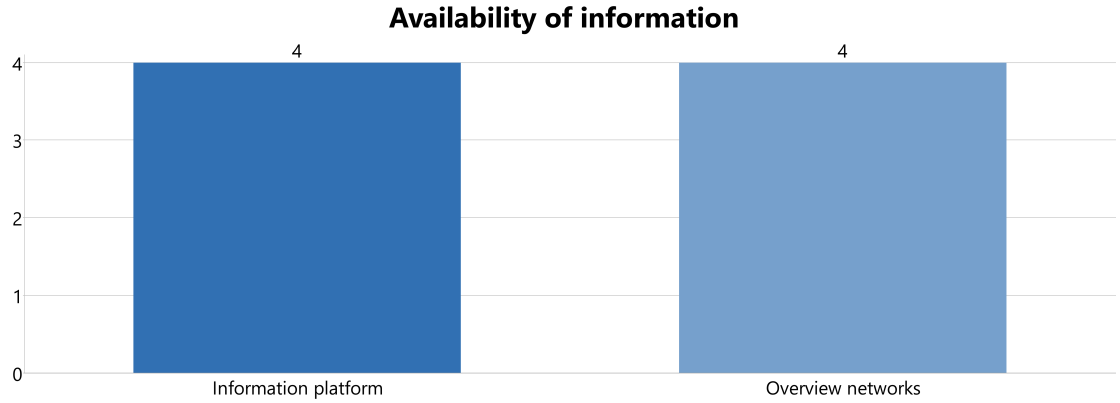


Figure 10: Subcode statistic availability of information

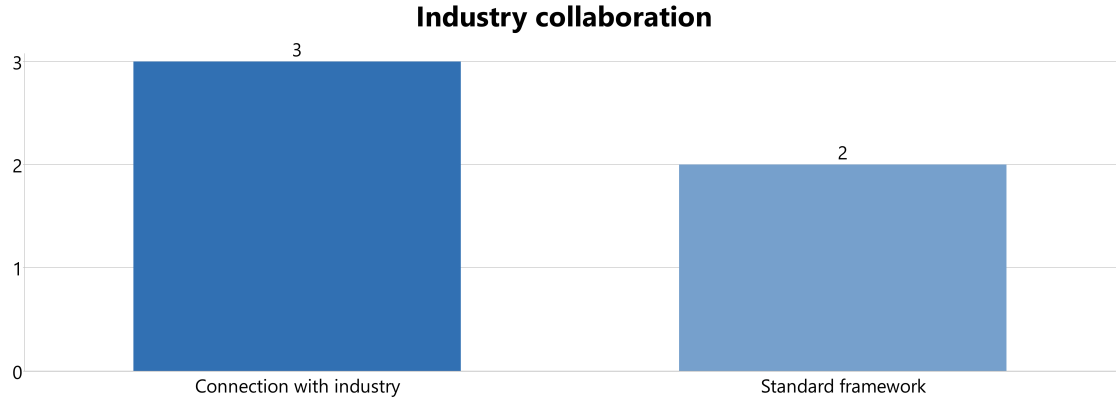


Figure 11: Subcode statistic industry collaboration

of Industry and Commerce (Industrie- und Handelskammer, IHK), which is in part financed by company contributions, could create a portal and perhaps provide low threshold funding for increased exchange. E12 highlights the need for more initiatives promoting structured cooperation projects between start-ups and industry:

> "We need functioning projects with pharmaceutical companies and healthcare actors, who really commit to carrying out joint projects with start-ups in a certain time frame including certain payments." (E12)

In addition, E9 advocates more mentorship from and collaboration with industry in the very early stage. He especially sees improvement potential in the collaboration with healthcare professionals as they are chronically overworked. The interviewee also suggests expanding programs such as Innovate Healthcare to foster communication between students and the respective clinics and making them available at an earlier stage.

Lastly, E6 and E15 suggest the implementation of a standard framework for the cooperation of start-ups and the health sector. For founders it can be challenging to connect with healthcare professionals to test their solutions, as many healthcare providers have reservations due to regulatory or compliance reasons or other concerns. However, the requirements are often the same:

> "Actually, the questions are always the same. People have to get an interface to the data and the hospital infrastructure to try things out. And meet with relevant doctors, who are sensitized to the fact that they can help to drive innovation in their daily hospital routine." (E15)

Therefore, the two experts propose a standard intellectual property deal to establish cooperation between large health care providers and young researchers and founders by lowering the barrier for clinics to engage in this type of partnership. As an example, E15 mentions the office of technology licensing in Stanford, which facilitates the collaboration with clinics and the distribution of royalties.

5.3. Alternative ecosystem framework

Two interviewees mentioned potential approaches for an alternative framework of the entrepreneurial ecosystem elements. The framework described by E14 is illustrated in Figure 12. According to the interviewee, formal institutions, in the sense of educational and research institutes, and talent

form one side of the framework. These two elements correlate with each other as talent emerges from the institutions. On the other side, he positions demand and finance, where ideas can be tested and are supported by financial resources. Networks form the connection between the two sides. The networks are also responsible for creating the culture which surrounds the entire ecosystem and ensuring knowledge is exchanged between the different elements and stakeholders. Physical infrastructure would be considered as flanking element, which does not work in isolation, but supports the ecosystem. E1 adds that for him the networks are a combination of intermediaries, knowledge, leadership, talent, and physical infrastructure and that all these factors together constitute a strong network.

6. Discussion

In the following, the research findings are discussed in the light of the research question and the current literature in the field of entrepreneurial ecosystems. Following this, the methodological approach is critically reflected, and research limitations are discussed.

6.1. Discussion of the results

The aim of this study was to qualitatively evaluate the Munich entrepreneurial ecosystem in the health sector to gain an understanding of the current state of the different elements and develop recommendations to advance the ecosystem. Therefore, the central research question was defined as "How can the Munich entrepreneurial ecosystem in the health field become more efficient?". In the following, the main research findings from the expert interviews are discussed, taking into consideration the theoretical concepts of entrepreneurial ecosystems and the present research question.

With regards to the current life cycle stage, the Munich ecosystem is likely still in its growth phase. This phase is marked by an increasing number of new start-ups, support programs and financing options and talent becoming more entrepreneurially minded.[73] The results of the interviews confirm this proposition. First, it was mentioned multiple times by several interviewees that there is a mindset shift among students, academics, and industry professionals towards the start-up sector and entrepreneurial projects. Second, specialized support structures such as intermediaries and networks are increasingly available for the health sector. Multiple experts commented that there is still an insufficient number of networks in Munich, but they are noticing constant improvements. For instance, new programs such as the TUM Venture Lab Healthcare are being formed. And third, the leadership feedback loop into the ecosystem is not fully established yet, as there are not enough experienced founders yet in Munich. Multiple experts think that this is a function of time and are confident that the amount of leadership and feedback to the ecosystem will improve in the next years, as more companies in the sector either become successful or fail.

The interviews revealed the perceived quality of each element and improvement ideas, which were grouped into five key areas. According to the literature on ecosystem governance, it is important to particularly focus on the weaknesses of the ecosystem, as they can form bottlenecks and obstruct further development and growth.[74] Therefore, the improvement ideas will be discussed in order of the corresponding ecosystem elements, starting with the lowest scored element. Physical infrastructure and leadership were rated as the weakest parts of the Munich ecosystem. In terms of the infrastructure, most experts criticized that affordable laboratory and office space is scarce in Munich. The available incubators often have long wait times and tend to favor later stage companies. To improve this weakness, several experts suggested the creation of more shared lab space within incubators, where several teams share equipment and office space. Not only would this option be more affordable for very early-stage start-ups, but the resulting exchange between the entrepreneurs has been claimed to be highly educational. Leadership has been rated equally low by the interviewees. On one hand, there are simply not enough successful founders in the health sector in Munich. On the other hand, the few existing leaders are often not available as mentors in the ecosystem. Leadership is a function of time and the consequence of a productive ecosystem. Therefore, to increase leadership, the ecosystem as a whole must be supported so it can generate productive entrepreneurship and ultimately valuable leadership. However, there has been one actionable improvement idea, namely the creation of "Entrepreneurial Bafög", which combines early-stage financing of students and young professionals with the creation of more mentorship and entrepreneurship examples in Munich. Obliging funded students to share their entrepreneurship experience would increase the availability of mentorship in the ecosystem. In accordance with governance principles in ecosystem literature, these two factors should be given the highest priority when designing governance initiatives.

The element of formal institutions, including the regulatory framework and public initiatives, is the third of the low score elements. According to Corrente et al. (2019), "Government Programs" are the second most important factor determining the success of ecosystems, as they can significantly accelerate or inhibit the growth of start-up companies.[75] The interviewees often criticized the highly bureaucratic processes and the lack of tax incentives for start-ups. Furthermore, the incentive structure of public fund committees has been claimed to be inefficient. In terms of improvement ideas, several experts mentioned the need for financial incentives for entrepreneurs, such as simplification of the tax

[73] See Cantner et al. (2021, pp. 413-417).

[74] See Audretsch et al. (2016, p. 373).

[75] See Corrente et al. (2019, p. 513).

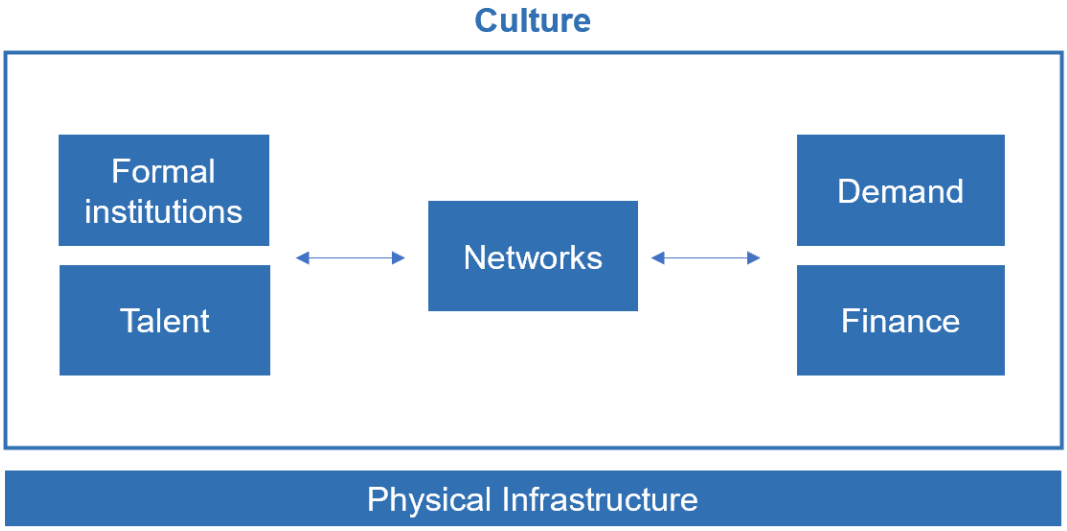


Figure 12: Alternative entrepreneurial ecosystem framework

on start-up stock options or discounted office space. Isenberg (2010) also emphasizes the need to reform the regulatory framework to enable productive entrepreneurship. This can include removing administrative and legal barriers and setting the right financial incentives.[76] The interviewees also suggested linking public support programs more closely to the private sector by creating more balanced committees and incentivizing decision makers with carried interest. This suggestion corresponds to the third principle of Isenberg (2010), which states that the public sector should be engaged in designing entrepreneurship policies and programs.[77] Although reforming formal institutions is a difficult and lengthy process, it can significantly contribute to accelerating entrepreneurship. However, this approach cannot be effective in absence of the other improvements mentioned in this section.

The availability of finance was rated mostly positively by the interviewees, stating that Munich is home to a good network of VCs and other private investors. Still, the most frequently mentioned improvement point for the Munich ecosystem was to increase early-stage financing and enable entrepreneurs to obtain small grants in an uncomplicated manner. These ideas include diversifying the current support programs to enable life science start-ups to be funded for longer time periods and creating programs which supply an initial low amount of funding, followed by potential successive rounds. It is surprising that no suggestions were expressed regarding the main point of criticism, which is the lack of later stage financing options in Germany. However, it can be argued that counteracting this problem is rather difficult since this is mainly caused by the lack of large stock exchanges and late-stage investors.

The entrepreneurial culture was described to be very strong within networks, but not as pronounced in broader society. Corrente et al. (2019) declared culture and the associated social norms as the most important factor impacting the success of entrepreneurial ecosystems.[78] In line with these findings, six interviewees proposed measures to increase entrepreneurial education to improve the entrepreneurship culture in Munich. Since natural scientists, engineers and medical doctors are the most qualified to start a business within the health sector, it is important to include optional entrepreneurship courses in their studies. In addition, promoting entrepreneurial education before university can not only create this understanding for entrepreneurship earlier, but also sharpen the awareness in other professions in the sector, such as nurses and technical assistants. Finally, it is suggested by the experts to highlight successful companies in the health sector to improve the culture. These ideas match two of Isenberg's principles "Get a big win on the board" and "Tackle cultural change head-on". He emphasizes that visible, successful ventures can reduce the public perception of risks and barriers associated with entrepreneurship.[79] These initiatives are relatively cheap and easily implemented compared to other suggestions like financial incentives and incubator development. In addition, a good entrepreneurial culture is widely regarded to be of paramount importance for a successful ecosystem and should therefore be prioritized within governance initiatives.

Several improvement ideas relate to the entrepreneurial networks in Munich, which have been given an intermediate score. The principles of entrepreneurial governance contend that entrepreneurs should act as central leaders within the stakeholder network of the ecosystem.[80] This is reflected by

[76] See Isenberg (2010, p. 49).
[77] See Isenberg (2010, p. 44).
[78] See Corrente et al. (2019, p. 513).
[79] See Isenberg (2010, p. 47).
[80] See Kuckertz (2019, p. 479f).

the desire of the interviewees to see more entrepreneurs in key positions of networks and incubators. However, this is directly dependent on the presence of successful founders and leadership. One relatively straightforward improvement for networks is to create an overview of the existing networks and support programs, for instance by informing founders when they are registering their business. This simple measure would help to increase the reach of current programs to more start-ups.

The elements of talent and intermediaries have been generally rated very positively. Owing to the two major universities TUM and LMU, there is a lot of skilled talent in Munich compared to other German cities. Intermediaries are also sufficiently available and are becoming more specialized towards the health sector. With regards to knowledge, four interviewees suggested the creation of an information platform to share information regarding practicalities such as taxes, cooperation partners and hiring, as all start-ups encounter similar challenges. Lastly, demand was the highest rated element in the Munich ecosystem. Nevertheless, one of the five improvement areas relates to this factor. Despite the presence of many large companies in Munich, they must be made more accessible to start-ups. This should include stronger matchmaking by existing institutions such as the IHK. Moreover, two experts advocated a standard framework for the collaboration with hospitals and healthcare professionals. Although the healthcare sector is currently already overburdened, the one-time investment in such a framework could reduce the required time for future projects, which could in turn lead to more innovation in the health sector.

The entrepreneurial ecosystems theory was frequently critiqued in recent literature for its tautological concept and its missing explanation of cause and effect. The same points were raised in several interviews, and two experts even outlined an alternative approach to the ecosystem framework. The suggested framework by E14 addresses the criticized weaknesses of the framework, as it starts to clarify relationships and interdependencies between the elements. Furthermore, the approach differentiates between critical elements, such as talent, culture, networks, demand and finance, and flanking means such as physical infrastructure. Although it can be argued that physical infrastructure is also essential in building an ecosystem as it provides the space for entrepreneurship and innovation to occur, it cannot be allocated the same weight as the central actors of the ecosystem, such as talent, and rather functions as enabler.

The elements leadership, intermediaries, and formal institutions in terms of a regulatory framework were not yet clearly positioned in the framework by E14. E1 mentioned that intermediaries and leadership should form part of the networks, and by that logic be placed in the center of the framework. It can be argued that leadership strongly correlates with networks, as these are often shaped and operated by leaders of the ecosystem. Intermediary services could either form part of networks or act as ecosystem enabler, supporting ventures whenever necessary. In my opinion, the regulatory framework can be positioned next to the physical infrastructure as an ecosystem enabling element. Regulations such as taxation are flanking factors, which guide and regulate entrepreneurship, but should not act as a central force. Although this alternative framework is by no means complete, it can provide a starting point for further refinement to ultimately arrive at a structure which clarifies the cause and effect between the elements and can therefore provide some insight to inform further research and ecosystem policy.

6.2. Discussion of the methodological approach

The qualitative approach to the research question had the advantage that, despite the relatively low number of participants, a lot of opinions on the Munich ecosystem and suggestions for improvement could be acquired. The experts were purposefully not guided in any direction, which resulted in honest reviews of the current strengths and weaknesses of the ecosystem and a variety of improvement areas and concrete recommendations.

The selection of interview partners focused on founders, investors and managing directors of support organizations in the wider health field. However, the associated fields of biotech, medtech and digital health are very different in their resource requirements, product development cycles and customers. For example, while life science start-ups require laboratory space with elaborate equipment and materials, digital health start-ups simply require a small office space and computers. Their products and customers are also very different. While product development cycles in life science and medtech typically take years to complete and require extensive clinical trials and subsequent regulatory approval, many digital health products can be developed significantly faster and tested more easily and on a broader audience. These differences complicated the generation of main improvement areas applicable to the entire health sector.

By only selecting entrepreneurs, investors and support organizations, the measurement of the entrepreneurial ecosystem parameters was potentially slightly biased. This study neglected other stakeholders of the entrepreneurial ecosystem such as customers in larger corporations and start-up employees. Therefore, especially the culture parameter might have an upward bias, as the interview participants tend to be surrounded by entrepreneurial networks and support groups with a strong entrepreneurial mindset. As described in the results section, many companies and healthcare providers are perceived to have certain reservations about the cooperation with young start-ups and talent oftentimes prefers the working conditions in larger corporations. However, this study aims to reflect the thoughts and opinions of the central stakeholders of the ecosystem and does not intend to provide an objective measurement. In addition, the selection of experts was aligned with a similar study in the research field by Spigel (2017), who conducted a case study on two ecosystems in Canada.[81]

[81] See Spigel (2017, p. 59).

To structure the analysis and the interview guideline, the entrepreneurial ecosystems framework was used. Although the framework was helpful in organizing the elements and provided a structure for the expert interviews, it has several shortcomings. First, although the elements are collectively exhaustive, which the experts all agreed on, they are not entirely mutually exclusive. This makes it extremely difficult to analyze the elements individually and understand their role in the ecosystem. For instance, formal institutions, talent, and knowledge strongly correlate with each other. Formal institutions such as good universities form the basic condition for well-educated talent. Knowledge, on the other hand, is often tied to chairs and professorships at universities and to the talent conducting the research or operating a company. Intermediaries and formal institutions are also not always selective, as networks and support programs by formal institutions could be allocated to both elements. For instance, TUM Start-up Consulting is a consultancy service by the university to support students or scientists at TUM who want to start their own company and could be attributed either to the formal institutions or intermediaries.

In addition, with the present selection of interviewees, not every element was applicable to every participant. For instance, the element of finance was difficult to rate for founders who have grown their company organically and therefore never obtained external financing through public grants or private investors. The elements are also not all easily applicable to one single location. In terms of intermediaries, it is not important that all necessary intermediate services like consultancy services and legal advice are situated in Munich. Many interviewees utilize services and collaborate with consultants, who are based in Berlin or even outside of Germany. Demand is also not constrained to one location, as the health sector is a global industry. Furthermore, the framework assumes that all ten elements are equally important in an entrepreneurial ecosystem. However, it can be easily argued that certain elements should be given a stronger weighting, such as demand and finance, without which a company could not sustain itself. Finally, the results obtained were limited by the defined sample size. Theoretically, it would be optimal to conduct expert interviews until no new insights can be gained. Unfortunately, due to time restrictions this was not possible.

As part of the analysis, the rating of the individual elements by the interviewees was illustrated in a box and whisker plot, including the calculation of the mean and quartiles for each element. Despite the primarily qualitative form of analysis, this form of presentation seemed appropriate to give an impression of the perceived quality of the different elements. Due to the small sample size of 15, these numbers are not generalizable since they merely express the opinions of the selection of experts and are therefore purposefully not mentioned frequently in the rest of the thesis.

7. Conclusion

The objective of this study was to qualitatively assess the entrepreneurial ecosystem in Munich and generate recommendations for improvement, answering the research question “How can the Munich entrepreneurial ecosystem in the health sector become more efficient?”. The aim of the thesis was achieved by extracting suggestions from the expert interviews and subsequently organizing and summarizing them into five main improvement areas. First, financial support could be improved by offering more early-stage financing, entrepreneurship incentives as well as updating the structure of public funds. Second, incubators could be refined to offer more shared laboratory space and batches, and networks and mentorship need to be improved, for instance by incorporating more founders. Third, entrepreneurial education should be increased both before and during university and successful ventures need to be highlighted more. Fourth, the availability of information needs to be improved for founders. This includes creating an overview of the available networks and a platform with information about practicalities in the entrepreneurship process. Finally, collaboration with industry needs to be simplified, by creating a stronger matchmaking process and generating a standard framework for cooperation projects.

These findings are not only valuable for the study of entrepreneurial ecosystems, but also have practical applications. The research community can benefit from the alternative ecosystems framework to enhance the ecosystems framework. Incorporating the internal dynamics and relationships within an ecosystem is essential for a more nuanced understanding of how entrepreneurial ecosystems result in productive entrepreneurship. Perhaps even more importantly, the results offer an assessment of the entrepreneurial ecosystem in the Munich health sector and potential areas for improvement. As stated in the literature review, governance initiatives often fail to accomplish the desired results, as strategies are simply copied from other locations.[82] Therefore, it is crucial to focus on local characteristics to devise governance initiatives aiming to improve the ecosystem. This case study on the Munich ecosystem identifies the local strengths and weaknesses and suggestions by central stakeholders of the ecosystem. Consequently, these results can be taken as a starting point for a strategy to make this ecosystem more effective. The results of this study might also lead to indications in similar research and knowledge intensive fields. Industries such as aviation and automotive have similarly long development periods and therefore similar ecosystem requirements.

Further research avenues in the field of entrepreneurial ecosystems could encompass the refinement of the framework and case studies in other geographic locations or with a broader selection of interviewees. As elaborated in the research limitations, the ecosystems framework is not entirely

[82] See Colombelli et al. (2019, pp. 505-507).

mutually exclusive and does not clarify the connections between the elements. For further research on entrepreneurial ecosystems, it could be helpful to further rethink the framework and incorporate the relationships and connections between the elements. In addition, the development of a dynamic perspective to explain the evolution of an ecosystem over time would be useful to arrive at a holistic understanding of entrepreneurial ecosystems. Secondly, this type of case study could be applied to other locations, for instance Berlin, to create an understanding of the quality of other ecosystems in the health sector. Further research could also look at the topic from a broader perspective, such as consulting a broader selection of interviewees. By including, for instance, employees of start-ups and industry professionals collaborating with entrepreneurs, a larger proportion of stakeholders would be covered which might offer a more realistic understanding of the ecosystem.

The entrepreneurial ecosystem in the health sector in Munich has gained considerable momentum and is currently growing in size. The related biotech, medtech and digital health industries generate a real impact for our economy, society, and technological expertise. Therefore, it should be of utmost importance to policymakers, investors, and other ecosystem stakeholders to support and further expand this entrepreneurial ecosystem.

References

Alvedalen, J., & Boschma, R. (2017). A critical review of entrepreneurial ecosystems research: towards a future research agenda. *European Planning Studies*, *25*(6), 887–903.

Audretsch, D. B., Cunningham, J. A., Kuratko, D. F., Lehmann, E. E., & Menter, M. (2019). Entrepreneurial ecosystems: Economic, technological, and societal impacts. *The Journal of Technology Transfer*, *44*(2), 313–325.

Audretsch, D. B., Lehmann, E. E., & Menter, M. (2016). Public cluster policy and new venture creation. *Economia E Politica Industriale*, *43*(4), 357–381.

Baeyens, K., Vanacker, T., & Manigart, S. (2006). Venture capitalists' selection process: the case of biotechnology proposals. *International Journal of Technology Management*, *34*(1/2), 28–46.

BfArM. (2020). The Fast-Track Process for Digital Health Applications (DiGA) according to Section 139e SGB V.

Cantner, U., Cunningham, J. A., Lehmann, E. E., & Menter, M. (2021). Entrepreneurial ecosystems: a dynamic lifecycle model. *Small Business Economics*, *57*(1), 407–423.

Colombelli, A., Paolucci, E., & Ughetto, E. (2019). Hierarchical and relational governance and the life cycle of entrepreneurial ecosystems. *Small Business Economics*, *52*(2), 505–521.

Colombo, M. G., Dagnino, G. B., Lehmann, E. E., & Salmador, M. (2019). The governance of entrepreneurial ecosystems. *Small Business Economics*, *52*(2), 419–428.

Corrente, S., Greco, S., Nicotra, M., Romano, M., & Schillaci, C. E. (2019). Evaluating and comparing entrepreneurial ecosystems using SMAA and SMAA-S. *The Journal of Technology Transfer*, *44*(2), 485–519.

Dang, A., Arora, D., & Rane, P. (2020). Role of digital therapeutics and the changing future of healthcare. *Journal of Family Medicine and Primary Care*, *9*(5), 2207–2213.

Deloitte Centre for Health Solutions. (2015). *The challenge of compliance in life sciences: Moving from cost to value*.

Flick, U. (2019). Gütekriterien qualitativer Sozialforschung. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (pp. 473–488). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-21308-4_33

Germany Trade & Invest. (2018a). *German Digital Health Industry Profile*.

Germany Trade & Invest. (2018b). *German Medical Biotechnology Industry Profile*.

Grand View Research. (2022). Digital Health Market Size, Share & Trends Analysis Report By Technology (Healthcare Analytics, mHealth), By Component (Software, Services), By Region, And Segment Forecasts, 2022 - 2030.

Isenberg, D. (2010). How to start an entrepreneurial revolution. *Harvard Business Review*, *88*(6), 41–50.

Kaiser, R. (2014). *Qualitative Experteninterviews* (2nd).

King, N., Horrocks, C., & Brooks, J. (2019). *Interviewing in qualitative research* (2nd).

Kuckartz, U. (2020). *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung* (5th).

Kuckertz, A. (2019). Let's take the entrepreneurial ecosystem metaphor seriously! *Journal of Business Venturing Insights*, *11*, 476–494.

Lechner, C., & Dowling, M. (1999). The Evolution of Industrial Districts and Regional Networks: The Case of the Biotechnology Region Munich/Martinsried. *Journal of Management and Governance*, *3*(4), 309–338.

Leendertse, J., Schrijvers, M., & Stam, E. (2021). Measure Twice, Cut Once: Entrepreneurial Ecosystem Metrics. *Research Policy*, 476–494.

Mack, E., & Mayer, H. (2016). The evolutionary dynamics of entrepreneurial ecosystems. *Urban Studies*, *53*(10), 2118–2133.

Maresova, P., Penhaker, M., Selamat, A., & Kuca, K. (2015). The potential of medical device industry in technological and economical context. *Therapeutics and Clinical Risk Management*, *11*, 1505–1514.

Mayring, P. (2016). *Einführung in die qualitative Sozialforschung: Eine Anleitung zum qualitativen Denken* (6th).

Misoch, S. (2015). *Qualitative Interviews* (1st).

NVCA. (2021). National Venture Capital Association Yearbook.

Price Water House Coopers. (2009). Drug Discovery and Biotechnology in Germany.

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.

Senior, M. (2021). Biotech bubbles during the global recession. *Nature Biotechnology*, *39*(4), 408–413.

Spigel, B. (2017). The Relational Organization of Entrepreneurial Ecosystems. *Entrepreneurship Theory and Practice*, *41*(1), 49–72.

Spigel, B., & Harrison, R. (2018). Toward a process theory of entrepreneurial ecosystems. *Strategic Entrepreneurship Journal*, *12*(1), 151–168.

Stam, E. (2015). Entrepreneurial Ecosystems and Regional Policy: A Sympathetic Critique. *European Planning Studies*, *23*(9), 1759–1769.

Stam, E., & Spigel, B. (2018). Entrepreneurial Ecosystems. In *The SAGE Handbook of Small Business and Entrepreneurship* (pp. 407–421). SAGE Publications Ltd. https://doi.org/10.4135/9781473984080.n21

Stam, E., & van de Ven, A. (2021). Entrepreneurial ecosystem elements. *Small Business Economics*, *56*(2), 809–832.

van Norman, G. (2016). Drugs, Devices, and the FDA: Part 2: An Overview of Approval Processes: Fda Approval of Medical Devices. *JACC. Basic to Translational Science*, *1*(4), 277–287.