JUMS

Volume 8, Issue 3, September 2023

JUNIOR MANAGEMENT SCIENCE

Published by Junior Management Science e.V.

Junior Management Science 8(3) (2023) 569-590

Junior Management Science

journal homepage: www.jums.academy

The more, the better? Diversification Trends in Executive and Supervisory Boards in Germany and their Potential Effects

Bogdan Bedelev

Ludwig-Maximilians-Universität München

Abstract

In 2015, Germany passed the Gender Quota Law, and while some countries compelled listed companies to reserve at least 30% of their executive seats for women, imposing fines on the firms that failed to comply, Germany favoured soft-law quotas with almost no penalties. Additionally, this policy focused solely on supervisory board quotas and measures to counteract women's under-representation, neglecting other demographic and cognitive groups. Given the increasingly diverse population in Germany, it is necessary to study the role of other diversity dimensions in the board composition from the financial and social perspectives and whether there are any development trends in the German boards. In my Bachelor thesis, I study whether there are any diversity improvements in the composition of German-listed companies' executive and supervisory boards, presenting recent academic findings on the drivers and the effects of diverse boardrooms. Moreover, I conduct a descriptive analysis of the German board diversity trends, implementing a novel diversity index of Bernile, Bhagwat, and Yonker (2018) covering various diversity facets.

Keywords: Controlling; Leadership; Board diversity; ESG; Corporate governance.

1. Introduction

It is widely believed that the Great Depression of 2008 resulted from governance failures (Berglof, 2011, p.500), as corporate boards have been condemned for the inability to impede the critical period and prevent severe economic fallouts. Many economists like Guest (2019) attribute this failure to the lack of diversity in the companies' boards, initiating a new wave of analyses and an extended appeal for diversity. In academic and regulatory spheres, board characteristics, such as gender, ethnicity, and functional background, have gained growing attention, as these might presumably influence the effectiveness of the decision-making process (Fernández-Temprano & Tejerina-Gaite, 2020, p.325).

Additionally, as societies are becoming more diverse in the last years and companies rely increasingly more on cross-functional teams to address complex issues, the question of how workgroup diversity affects team performance is more relevant than ever before (Plaut, 2010). The rapid technological change and globalisation have led to unprecedented environmental competitiveness, and thus, many proclaim diversity as a potential mean to develop more responsive and adaptive companies (Wright & Snell, 1999, p.49).

Acknowledging the substance of this topic, many researchers and practitioners have partaken initiatives during the last decade to generate a broader understanding of the diversity effects and its necessity. The present thesis tries to contribute to this mission as it examines the existence of the diversity trends in the German corporate context and the potential economic effects of board diversity, exploring the emerging literature on this issue.

To start with, I outline the primary goals of corporate governance regulations and their relation to board diversity. The focus lies primarily on the German corporate governance system and the enacted gender quota regulation, as this sets the ground for the later descriptive analysis. Then,

First and foremost, I would like to thank Prof. Dr Christian Hofmann, Prof. Dr Nina Schwaiger, Martin Viehweger, and the whole Institute for Accounting and Control team at LMU Munich for their warmest support, valuable feedback, impeccable supervision, and inspiring ideas throughout the time of conducting my Bachelor thesis. Further, I would love to thank my family and my dearest friends, especially Georgios Kantzis, who always supported me during the process of my research and motivated me to move forward and dream big!

DOI: https://doi.org/10.5282/jums/v8i3pp569-590

I describe the main idea of board diversity, its dimensions, and the several possible methods to classify these characteristics, as this might be valuable when conducting scholarly research. However, during the study, my attention is mainly attributed to the demographic and cognitive distinction of diversity, as these two groups have been focal points for most researchers (Bernile et al., 2018; Erhardt, Werbel, & Shrader, 2003; Kilduff, Angelmar, & Mehra, 2000; Williams & O'Reilly III, 1998).

After these clarifications, I move to the central part of this thesis. First of all, it is essential to delineate potential determinants of board diversity. In *Section 2.4.1*, I concentrate on the firm internal drivers, these being firm and board size, the degree of shareholder concentration, as well as the interconnected influence of different diversity dimensions. Furthermore, since all firms are linked to their external environment (Daft, 2013), *Section 2.4.2* presents conceivable diversity driving factors outside the firm. It is worth noting that the firm external environment could play a principal role in this study, as most arguments regarding the effects of board diversity rely to some extent on this component.

The attention then turns to the analysis of the possible effects of board diversity. In *Section 2.5.1*, I summarise the results from the literature on the board diversity impacts on firm performance. Then, I focus on the correlation between firm risk and board diversity, as it could be a critical area for the profitability and the existence of many companies. This is followed by a discussion over the effects of board diversity on board compensation. Lastly, I provide an overview of other potential effects, concentrating on corporate policy formulation and innovation processes.

Moreover, even though various sources yearly document the board diversity developments worldwide, especially after the introduction of the gender quotas in many countries, these reports commonly focus on female representation. This fact could explain my interest in the descriptive research in *Section 3*, where I describe the measurement methods and then present actual evidence on the board diversity trends in German listed companies over the last twenty years. The uniqueness of this study is the implementation of a novel diversity index following Bernile et al. (2018), which aims to simultaneously capture the development of different diversity elements, such as gender, age, nationality, university affiliation, financial expertise, and board tenure of individuals. Further, I also disaggregate this diversity index and discern the development of the boards' demographic and cognitive traits and then discover the tendencies of each particular diversity attribute. In the last sections, I discuss the results, build several propositions for future research that stem from the evident trends and insights on the diversity effects, and conclude my thesis, also mentioning its limitations.

2. Literature Review on Board Diversity

2.1. Definition of Corporate Governance

The first review of the corporate governance topic presumably dates back to the times when this concept's definition did not exist. Berle and Means (1932) brought up a problem in their book when a firm manager did not act in line with its owner's interests, referring to one of the fundamental concerns of upper-management today, the so-called *principal-agent problem* (Eisenhardt, 1989). By delegating the decision-making authority from the *principal* (shareholders) to the *agents* (managers), an agency problem can occur due to the separation of ownership and control, affecting the wealth of both parties (Jensen & Meckling, 1976, p.309). Initially, corporate governance practices appeared to reduce such interest conflicts, but today, they have evolved into a more multifaceted topic, referring to the pool of mechanisms that influence the decision-making process of managers in a firm, assure that they pursue the objectives determined by the shareholders (Campbell & Mínguez-Vera, 2008, pp.436-439), and specify each member's rights and responsibilities (OECD, 2005).

Nevertheless, there are notable differences in corporate governance structures worldwide. While in the Anglo-American system, the prime objective is the optimal introduction of incentives and control to maximise the return on equity (ROE) (Shleifer & Vishny, 1997), in other countries, like Germany, the regulations strive to contemplate a broader spectrum of interests, such as these of the firm's employees and customers, and the potential conflicts with each other (Schmidt & Tyrell, 1997, p.344). Regardless of the corporate governance structure, a supervisory board of directors belongs to the essential firm internal mechanisms dedicated to ensuring that the shareholders' and managers' interests are closely aligned, to determining the overall corporate strategy, and to selecting, rewarding, or disciplining incompetent managers (Fauver & Fuerst, 2006, p.675).

Since the composition of the boards is seen as a powerful tool to improve corporate governance standards, today, many Corporate Governance Codes primarily concentrate on issues such as board diversity, board size, and the independence of directors (Carter, D'Souza, Simkins, & Simpson, 2010; Carter, Simkins, & Simpson, 2003). The existence of the diversity topic in the regulatory agenda is relevant due to its broad economic impacts, including not only the possible enhancement of the independence and the monitoring abilities of corporate boards, the generation of fresh ideas and perspectives (Anderson, Reeb, Upadhyay, & Zhao, 2011; Baranchuk & Dybvig, 2009), explained in *Section 2.5*, but also the promotion of social equity as well as the equal opportunities' provision (Sarhan, Ntim, & Al-Najjar, 2019, p.762). Thus, appointing members who improve the board diversity could establish more inclusive and fair business structures, benefitting existing shareholders (Terjesen, Sealy, & Singh, 2009, p.320).

2.2. Corporate Governance and Gender Diversity Requirements in Germany

Examining the German corporate landscape, its governance system can be defined as a *coordinated market system* that provides more strategic relations between firms and their stakeholders (Lane, 2003). According to domestic law, German stock companies, such as companies or partnerships lim-

ited by shares and limited liability companies, must possess a two-tier board structure with personal separation, meaning that nobody can be a member of both boards in the same firm simultaneously (AktG, 1965; Dittmann, Maug, & Schneider, 2010, p.41). Generally, this structure strives to separate decision management and control (van Veen & Elbertsen, 2008, pp.388-389) and formalise the particular governance function of outside directors as representatives of the firm's shareholders (Fauver & Fuerst, 2006, p.675). Similar to American boards, the German *supervisory board* is responsible for long-term planning and monitoring regarding the fulfilment of the company's financial goals and the appointment, performance review, and salary determination of the second board members, known as the *executive board*. The executives are in charge of the firm's everyday operations and must report back to the supervisory board on the overall performance (Fauver & Fuerst, 2006).

The corporate governance regulations in Germany also consider the topic of diversity. In 2015, the German government introduced its board-level gender quotas by the "Act on Equal Participation of Women and Men regarding Leadership Positions within the Sectors of Private Economy and Public Service". The first pillar of this act is the *Fixed Gender Quota*, which inducts a mandatory 30%-quota for the underrepresented gender in the German non-executive boards of organisations limited by shares. The predominant criterion is that the company is listed on a stock exchange and is subject to the German Codetermination Act. The second pillar is the *Individual Gender Quota*, which applies to all publicly traded companies that are not subject to the first quota, such as limited liability companies. As its title suggests, firms are free to determine their individual quotas; however, if the actual share of women is below 30% after establishing such individual quotas, the target share cannot be below the actual fraction. Non-compliance with the established quota regulations is sanctioned by empty board seats, meaning that the supervisory board is considered void if the election results do not abide by the government's requirements. This sanction persists until new elections provide results conforming to the law (BGBl, 2015).

2.3. Definition of Board Diversity

The present German Corporate Governance Code states "The Supervisory Board shall determine specific objectives (...) while taking the principal of *diversity* into account." (Regierungskommission, 2019, p.7). As it becomes evident from the wording, the regulation attends to the whole phenomenon of diversity without emphasising any particular feature. However, for the subsequent descriptive study of the board diversification trends in Germany, it is crucial to expound on the meaning of diversity and its various dimensions.

According to Harrison and Klein (2007, p.1200), the term *diversity* still lacks an explicit definition in academic literature. Yet, one suggested denotation of diversity is that it represents the occurrence of differences among members of a unit concerning a common property. In the corporate governance context, the concept of diversity relates to the board composition and the combination of different attributes and characteristics of its members that can interplay with board processes and decision-making (Van der Walt & Ingley, 2003, p.219). Some examples of these attributes are the director's age, gender, and professional background (Milliken & Martins, 1996). Hence, board diversity refers to the mixture of human, social, and intellectual capital that the board comprises collectively and draws upon engaging in its governance obligations (Van der Walt & Ingley, 2003).[1]

Besides, several categorisation methods of the various diversity dimensions have emerged over the years. One common practice is to differentiate between its demographic (observable) and cognitive (unobservable) attributes. The first group encompasses easily detectable directors' features, such as gender, ethnicity, and other demographic characteristics; while the unobservable attributes' group covers the directors' educational, functional, and occupational background, experience, perceptions, and values (Kilduff et al., 2000; Milliken & Martins, 1996). Other researchers suggest distinguishing between task-related and non-task-related (or relations-oriented) diversity dimensions (Adams, de Haan, Terjesen, & van Ees, 2015). Specifically, the task-oriented category encloses the directors' work-related capabilities to collect, process, and exchange information, negotiate and allocate resources (Pelled, Eisenhardt, & Xin, 1999). Contrarywise, the relations-oriented group includes attributes that evoke social cognitive processes, guiding team communication (Jackson, May, Whitney, Guzzo, & Salas, 1995, pp.216-219).

Finally, Pelled (1996) encourages the combination of the two categorisation methods in a matrix, where each attribute is simultaneously distinguished based on observability and task-relatedness. This two-fold distinction is helpful to understand whether specific observable or unobservable attributes contribute to the enhancement of the job-related skills of the board members. Thus, unobservable traits like the director's functional and educational background are usually highly task-related because of their strong association with job performance and expertise. Consequently, ensuring the representation of all four categories in a board could boost thoughtful and creative decision-making (Pelled, 1996; Williams & O'Reilly III, 1998).

[1] *Human capital* refers to the skills acquired by individuals from training and experience (Cambridge Dictionary, 2021). *Social capital* relates to the liaison among people working in the same group, promoting its efficient function (Bowles & Gintis, 2002, p.F419). *Intellectual capital* stems from mental processes that cultivate inputs for economic activity and value to its owners (Luthy, 1998, pp.3-4).

2.4. Determinants of Board Diversity

Before elucidating the potential effects of board diversity, it is essential to address its different drivers. Yet, one has to bear in mind that these two topics are interrelated, as shareholders may consider the potential gains and costs while selecting new board members to maximise the firm value (Arnegger, Hofmann, Pull, & Vetter, 2014, p.1111).

In the following sections, I separately observe firm internal and external drivers of diversity. The firm internal drivers can be directly regulated by the company, whereas the external drivers consider elements of the environment that influence the firm from the outside and are not under its control (Daft, 2013).

2.4.1. Firm Internal Drivers

First of all, Arnegger et al. (2014) consider the firm's size as an essential driver of board diversity, proclaiming that while firm size positively affects the directors' occupational background diversity, the diversification effect on the boardrooms' internationalisation is concave. These relations emanate from the benefits and costs of communication and conflicts. On the one side, the *resource dependence theory* (Pfeffer & Salancik, 1978) elucidates the benefits, pointing to the various resources such as expertise and communication channels that directors introduce to the board. In this sense, shareholders would prefer board heterogeneity to access numerous skills, knowledge, and linkages to necessary external contingencies generated from diverse occupational and national backgrounds (Hillman, Shropshire, & Cannella, 2007; Pfeffer & Salancik, 2003). Moreover, as the degree of complexity usually increases with firm size (Lawrence & Lorsch, 1967), shareholders could favour more diverse boards to better deal with the increased need for supervisory and administrative inputs (Bantel, 1993b). On the other side, despite the improvements in the decision-making process and the unique cognitive attitudes, occupational and international diversity also incur costs in terms of communication speed within the board due to the usage of jargon (Horwitz & Horwitz, 2007, p.992). According to the *social categorisation theory* (Tajfel, 1974) and the *similarity attraction theory* (Byrne, 1971), differences in values and attitudes, the decrease of trust and cooperation, and the rise of potential conflicts can lead to a general poorer performance. Overall, these trade-offs are central when shareholders appoint new directors, as they can explain their decision to keep internationalisation low when the firm size increases, avoiding the vast communication costs, and instead to increase occupational diversity due to the more bearable downsides (Arnegger et al., 2014).

Closely linked to the firm's size, another critical determinant of board diversity is the size of the boardroom (Zald, 1969). According to Chaganti, Mahajan, and Sharma (1985), larger boards offer a broader range of functions than only the usual services, purely by having more directors to spread around within the organisation. This can also boost board diversity more easily (Klein, 2002). Moreover, according to Sanders and Carpenter (1998, p.159), the board's size might reflect the complexity of the firm's external environment because of the increased need to react to changes swiftly and to enhance the information-processing capacity. Thus, larger boards necessitate diverse skills and perspectives that stem from board members with different traits to efficiently steer the company through business intricacies (Luoma & Goodstein, 1999).

Besides, the degree of shareholder concentration also affects board diversity. Specifically, the emergence of the *stakeholder theory* was prompted by the increasing need to consider a wider range of societal interests. One proxy for such interests is the presence of minority shareholders, or the degree of shareholder concentration measured by the percentage of shares held by the significant shareholders (Kang, Cheng, & Gray, 2007). Thus, by widening the domain of corporate governance beyond major shareholders to other stakeholders, such as suppliers and employees, board diversity can promote procedural justice due to the direct representation of different interests in the corporate decision-making (Luoma & Goodstein, 1999, p.554). Consequently, it is expected that a lower shareholder concentration results in a broadly represented board (Kang et al., 2007, p.198). This may help the company legitimise its activities to promote corporate social responsibility by introducing non-economic considerations, like environmental awareness and community involvement, into decision-making and by fostering more open governance processes that better assure the representation of the stakeholders' interests (Hillman, Keim, & Luce, 2001; Kang et al., 2007).

Finally, some diversity dimensions might also boost the existence of other diversity traits in the board. For example, it is documented that female supervisory and executive board members tend to also differ in their skills, experience, and age (Casteuble, Lepetit, & Tran, 2019, p.3). They are inclined to be younger than their male peers (Adams & Funk, 2012, p.229), have higher levels of education, and more international experiences (Singh, Terjesen, & Vinnicombe, 2008). Additionally, Hillman, Cannella, and Harris (2002, p.758) find that, in Fortune 1000 companies, female directors and other board members who enhance diversity tend to come from various, non-business backgrounds.

2.4.2. Firm External Drivers

From the firm's external perspective, the industry in which a company operates may also affect its board diversity level (Brammer, Millington, & Pavelin, 2007). For instance, supported by the *organisational contingency theory* (Galbraith, 1973), companies in some "non-masculine" sectors (e.g., service industries) can capitalise on the impacts of diversified boardrooms more efficiently because of the better market insights and the more significant interplay between employees and customers that emerge from diversification (Ali, Kulik, & Metz, 2011; Jackson, Schuler, & Rivero, 1989). Hyland and Marcellino (2002) assert that more than any other dimension, the number of women in the boardroom is correlated with the firm's industry. Hence, companies in healthcare or technology-related sectors are

more likely to employ female directors than organisations in technic-specific industries, such as oil, commodities, and construction (Harrigan, 1981, p.624; de Cabo, Gimeno, & Escot, 2011). The board's age diversity is also significantly influenced by industry since companies in consumer services are more likely to appoint directors from various age ranges. Specifically, these companies address customers of all ages, and so a variety of age groups in the boardroom can better speak for the consumers' interests (Kang et al., 2007, p.196). On the contrary, Adams and Ferreira (2007, 2009) argue that boards tend to be less gender and age heterogeneous in riskier environments and industries to enhance the boards' monitoring abilities, improve the reaction speed to external changes, and avoid conflicts and difficulties in the decision-making process.

Furthermore, changes in the business environment are often associated with adjustments in the overall corporate strategy (Hillman, Cannella, & Paetzold, 2000, p.242). Since the board of directors intervenes in strategy formulation, it is involved in any significant strategy change to adapt to the external environment (Tushman & Romanelli, 1985). As mentioned earlier, the resource dependence theory supports that each company must keep up with the changes in the external environment to succeed. Therefore, the board's composition in terms of demographic or task-related characteristics may necessitate strategic alterations to keep transaction costs low and reap the benefits of enhanced communication channels, facilitating the company's strategic change (Pfeffer & Salancik, 1978).

Finally, each country's social, political, and economic structures can determine the level of board diversification. Naming Norway and Iceland as examples, Terjesen and Singh (2008, p.58) argue that if a country strongly embraces female representation in legislative and senior positions, this might influence society's beliefs about the management qualifications of women, thereby enhancing gender diversity in the boards. Additionally, in such states, the question of the general equality in opportunities is more likely to be on the political agenda, meaning that boardrooms may depict greater diversity on further dimensions other than gender, too. Lastly, the new corporate governance rulings also define corporate diversity, since depending on the country, they either compel or suggest the improvement of the share of underrepresented board directors (Terjesen et al., 2009).

2.5. Effects of Board Diversity

2.5.1. Effects on Firm Performance

Generally, boardrooms have at least four crucial functions: monitoring managers, providing information and counsel to principals, monitoring compliance regulations, and linking the corporation to its external environment (Monks & Minow, 2004). One basic proposition in the literature is that the boards' composition and diversity might affect how boards fulfil these functions, which are vital to determine firm performance (Carter et al., 2010).

The existing theoretical framework on the effects of board diversity on firm performance draws on various perspectives. According to the *agency theory*, for example, since the board of directors is a critical tool to monitor managers and mitigate conflicts between them and the shareholders (Fama & Jensen, 1983), an appropriate diversity level could enhance its monitoring role (Kandel & Lazear, 1992). Yet, Carter et al. (2003, p.37) argue that the agency theory alone cannot predict the effect of diversity since diverse boards may be marginalised, negatively affecting the monitoring outcome and, thereby, firm performance.

The *resource dependence theory* also plays a central role when analysing the impact of board diversity (Carter et al., 2010). The directors' established linkages provide the board with legitimacy and communication channels, aiming to reduce its dependence on external factors (Pfeffer & Salancik, 1978). Board members, for instance, could expand their networks, which in turn might enable firms to benefit from the improved access to their constituents (Hillman et al., 2000, p.239). The *human capital theory* complements this perspective, as directors with different backgrounds, skills, and experience provide their unique human capital to the boardroom, potentially enhancing firm performance (Becker, 1976).

In contrast, boardroom diversity may also hamper firm performance. Based on the *similarity-attraction paradigm*, the society perceives other individuals as "outsiders" when they differ from the main group (Byrne, 1971). In such cases, people might be reluctant to share information with them, leading to an interpersonal breakdown (Adams, Hermalin, & Weisbach, 2010). Extrapolating these thoughts to boardrooms, the *social psychology theories* propound that diverse cognitive abilities and perspectives can generate conflicts among groups that are similar in other traits (Williams & O'Reilly III, 1998). This, in turn, is likely to impede the board's cohesiveness and communication, protract decision-making, and diminish firm performance (Westphal & Bednar, 2005).

Moving from theory to praxis, the board diversity effects on firm performance have also been evaluated in empirical frameworks. In such literature, the attention is mostly on gender diversity, possibly due to data availability and the recently enforced gender quota regulations (Carter et al., 2010, p.397). Despite the vast number of studies, the empirical results are not unanimous, as some researchers proclaim a positive effect of gender diversity on firm profitability (Erhardt et al., 2003), firm value (Carter et al., 2010; Gordini & Rancati, 2017), and monitoring efficiency (Adams & Ferreira, 2009); however, other studies show a negative connection between gender diversity and the firm's gross profit and ROE (Haslam, Ryan, Kulich, Trojanowski, & Atkins, 2010) or even no statistical significance (Farrell & Hersch, 2005; Rose, Munch-Madsen, & Funch, 2013; N. Smith, Smith, & Verner, 2006). Rose (2007, p.411) explains that the negative results can be caused due to the process of socialisation where the unconventional board members, such as female directors and other board "minorities", must first adopt the behaviour and norms of the regular board members, thereby delaying the firm's processes. However, some researchers still contend the positive influence of gender diversity on firm performance

using the *critical mass argument* (Arena et al., 2015), which asserts that a visible impact on financial performance is only possible with a certain number of individuals with different traits (Kanter, 1977).

As for the other demographic and cognitive diversity dimensions, similar binary results are observed. Different nationalities, age groups, et cetera might also have a positive (Carter et al., 2010), a negative (Hafsi & Turgut, 2013; Mahadeo, Soobaroyen, & Hanuman, 2012) or a statistically insignificant (Rose, 2007) impact on firm performance.

For instance, nationality diversity might cultivate a profusion of experience and knowledge of various economic and operational environments, which could intensify competitiveness, group dynamics, and the quality of corporate social responsibility (Johnson, Schnatterly, & Hill, 2013; Khan, Khan, & bin Saeed, 2019). It could also support the innovative solutions' formulation and the efficient solving of complex tasks, shaping profitability and general performance (Nielsen & Nielsen, 2013). Contrarywise, Delis, Gaganis, Hasan, and Pasiouras (2017) state that communication is usually facilitated if the group members share a common background, similar ideas, and perceptions. Thus, the increased cultural diversity in the boards might hamper the company's smooth functioning due to communication problems that arise from social or language barriers and the lack of a common past, as posited by the *social identity theory* (Dumas, Phillips, & Rothbard, 2013; K. G. Smith et al., 1994).

Concerning educational diversity, most researchers uphold a positive influence on firm performance, as the conglomeration of different educational levels prompts various knowledge, ideas, and viewpoints, possibly resulting in better decisions and, thus, corporate performance (Bantel, 1993a; Kim & Lim, 2010). Nevertheless, Milliken and Martins (1996) find a negative influence, as board members with distinct educational backgrounds might also perceive, process, and respond to the issues they confront differently, resulting in a greater possibility of cognitive conflicts that hamper the firm's efficiency.

Finally, Kim and Lim (2010) and Mahadeo et al. (2012) signal the importance of age diversity for firm performance, as they highlight possible synergies between the younger board members' productivity and the experience of older ones.

To conclude, the empirical results are mixed, and as suggested by Adams and Ferreira (2009), the impact of board diversity on firm performance is likely to be heterogeneous. While some large companies might benefit from the enhanced diversity because they have more complex structures and need intensive monitoring stemming from different experiences, other companies might be harmed from this overmonitoring and the slacked in-board communication.

2.5.2. Effects on Firm Risk

From a theoretic perspective, the *upper echelons theory* (Hambrick & Mason, 1984) implies that the characteristics of executives are reflected in the firm's business strategies and performance outcomes (Bertrand & Schoar, 2003), and the diversity of the individuals' traits signifies corporate risk-taking decisions (Sila, Gonzalez, & Hagendorff, 2016). On the one hand, the preferences, incentives, and beliefs of homogenous groups could result in more idiosyncratic arrangements, as these groups pull less scrutiny within the board. This lack of internal governance would manifest itself in more volatile firm outcomes, arguing in favour of greater diversity in the boardrooms (Bernile et al., 2018). On the other hand, as discussed above, diversity might also trigger conflicts and disturb the board's decision-making process, making consensus harder and outcomes, such as firm risk, more unpredictable (Arrow, 1951).

From an empirical viewpoint, the negative relation between diverse boards and firm risk relies on the argument that such boards can enhance their monitoring and advisory role, helping the firm reduce risk in uncertain environments. Bhat, Chen, Jebran, and Memon (2019) examine the effects of diversified boards on risk considering both the relations- and task-oriented diversity dimensions in Chinese firms. They suggest that task-related diversity in terms of education and tenure could positively impact performance and alleviate risk more efficiently than demographic diversity. For instance, directors having diverse cognitive characteristics could make more effective decisions, reducing the chance to make suboptimal investments (Webber & Donahue, 2001). This view is also supported by Adams et al. (2015), as task-oriented diversity leads to moderated decisions and discipline. Nevertheless, Bhat et al. (2019, p.282) also stress that in the long run, the relations-oriented diversity is also vital in reducing corporate risk, as getting familiar with each other, board members can minimise communication problems. Another study conducted by Bernile et al. (2018) provides similar outcomes, analysing the relationship between general diversity (including gender, ethnicity, financial expertise, et cetera) and firm risk measured by the annual volatility of daily stock returns. The authors argue that board diversity smooths decision-making and eliminates problems related to groupthink.[2] In contrast, Coles, Daniel, and Naveen (2014) stress that diversity might actually generate groupthink if many board members are co-opted or have lengthier tenures since this may hinder the board's monitoring aspect and leave risk-moderation uncontrolled. Finally, Hambrick, Cho, and Chen (1996) also remain sceptical regarding the power of diversity to moderate risk, as greater diversity might lead to longer decision-making processes and reduce the firm's reaction speed, especially when the external environment is already volatile.

Disaggregating board diversity into its distinct traits, a growing number of studies has analysed the effect of gender on risk. A common conclusion is that female directors prefer lower risk in the financial decision-making process (Adams & Ferreira, 2009) and disclose more environment-

[2] *Groupthink* arises within a group of people who desire agreement or conformity at any cost, resulting, however, in unreasonable or dysfunctional decision-making (McCauley, 1989, p.251).

related information to avoid litigation issues (Liao, Luo, & Tang, 2015). Moreover, women are more conservative during investment decisions (Bernasek & Shwiff, 2001) and tend to hold less risky investment portfolios (Halko, Kaustia, & Alanko, 2012). The same conclusions regarding female risk aversion are also supported by Chen, Gramlich, and Houser (2017), as female directors are more cautious about firm reputational risks associated with aggressive tax strategies and generally avoid risky and challenging situations. However, Adams and Funk (2012) provide some opposite evidence and document that female directors concentrate more on stimulation and less on security, conformity, and tradition, tending to over-monitor and make riskier decisions than their male peers. This, in turn, decreases shareholder value (Ahern & Dittmar, 2012) and generates more firm-specific risks (Farag & Mallin, 2018). Finally, Berger, Kick, and Schaeck (2014) note that a higher presence of female board members results in less liquidity and more leverage in the firm's portfolio; however, when other diversity dimensions are also more starkly represented, portfolio risk is fairly mediated. The latter effect is explained by the fact that heterogeneous directors bear diverse experiences, allocating more time to portfolio selection and thereby, reducing risk.

In addition, other dimensions might also influence the board's risk-taking practices. Conventional wisdom, supported with empirical evidence, suggests that risk-taking appetite decreases with an individual's age (Campbell, 2006). Precisely, older managers tend to avoid high leverage and capital expenditures and to advocate higher cash holdings (Peltomäki, Sihvonen, Swidler, & Vähämaa, 2020, p.26) – practices that are not always chosen by younger directors (Davidson, Xie, Xu, & Ning, 2007). Moreover, concerning educational diversity, (Graham & Harvey, 2001, p.233) indicate that executives with a higher academic degree tend to use more sophisticated valuation techniques to assess and possibly, reduce corporate risk. As for the board's financial expertise diversity, financial experts have arguably lower costs in acquiring information on the environment's complexity and the associated transaction risks (Harris & Raviv, 2008; Minton, Taillard, & Williamson, 2014, p.352). Hence, they can recognise unprofitable risks more easily and advise senior executives against accepting them, as well as identify risks beneficial to shareholders and encourage executives to pursue them, increasing the shareholders' residual claims (Acharya, 2010).

To sum up, referring to the *contingency theory* (Fiedler, 1967), the discrepancy in the empirical results can be partially explained by the variations in the organisational environment in which risk-taking is considered. This theory implies that there is no universal management procedure to run an organisation, and management styles tend to be contingent on the environment's properties. This is why numerous studies report substantial differences when examining various diversity types (Saeed, Mukarram, & Belghitar, 2021).

2.5.3. Effects on Board Compensation

Next to the monitoring role, the supervisory board also approves the most important corporate decisions, such as recruitments or the design of the executives' payment packages (Monks & Minow, 2004). Thus, the managers' salary is affected by the efficiency of the board's supervision (Finkelstein, Hambrick, & Cannella, 1996).

On the one hand, according to the *optimal contracting theory*, the level of board diversity might influence its effectiveness, thereby enhancing its steering role (Adams & Ferreira, 2009), constraining managers from expropriating the shareholders' wealth avoiding overpayments (Stulz, 1988; Sarhan et al., 2019, p.767). On the other hand, the *managerial power hypothesis* suggests that close negotiations between a "weak" board member and a "strong" executive might lead to an inefficient executive compensation contract, increasing the agency problem (Bebchuk & Fried, 2004). Thus, more diverse board members may be perceived as tokens (Hillman et al., 2007; Ntim, 2015, p.173), meaning that corporate executives could easily influence the board's decisions, especially those related to their compensation structure (Sarhan et al., 2019, p.767).

Looking at the different diversity aspects, Adams and Ferreira (2009) investigate the role of gender diversity on the CEO's pay. The authors note that directors in gender-diversified boards receive higher equity-based compensations that provide more performance incentives. However, they find no statistical evidence regarding the impact of gender diversity on executive compensation, probably owing to the lower representation of women in the studied firms' compensation committees. Yet, Lucas-Pérez, Mínguez-Vera, Baixauli-Soler, Martín-Ugedo, and Sánchez-Marín (2015) and Benkraiem, Hamrouni, Lakhal, and Toumi (2017) conclude that as the presence of female directors positively affects the board's functioning, gender diversity might also improve the CEO compensation packages' design. This argument supports the idea that heterogeneous boards offer alternative perspectives that can improve the firm's strategic decision-making, which also includes the payment schemes (Milliken & Martins, 1996).

Additionally, as regards nationality diversity, a study of Scandinavian firms by Oxelheim and Randøy (2005) argues that this dimension has a significant positive effect on the CEOs' compensation. They suggest that a foreign board member representing their country's legislation could improve the incentive structure of the top management. Consequently, executives may be exposed to a clash between different corporate governance cultures, and the reconciliation of these systems could pose new challenges for them. For instance, this may raise the need for a new corporate language (Oxelheim et al., 1998), new reporting requirements or new investor-related activities (Useem, 1998), raising their pay. This higher CEO compensation could be seen as a risk premium for their increased duties due to the board's internationalisation (Oxelheim & Wihlborg, 1997).

Moreover, a positive relationship between foreign direc-

tors and executive compensation is also reported by Randøy and Nielsen (2002). The authors explain this correlation within the Scandinavian corporate context, where compensations are relatively low, and as such, when foreign managers from higher-paying countries, like the UK or the US, are employed, their salaries are adjusted upwards. The researchers stress that this channel does not imply that foreign directors are less monitored regarding their performance and compensation; on the contrary, it suggests that such directors might have more connections and be better able to employ competent chief executives.

2.5.4. Other Possible Effects

Board diversity influences many other aspects of the company as well, which are closely linked to firm performance and risk. To begin with, Bernile et al. (2018) review whether corporate financial and investment policies depend on board diversity and whether board diversity influences the corporate innovation. The authors contend that the policies adopted by diversified boards may be more stable and last longer. As discussed above, board diversity can shape firm volatility, meaning that policy corrections could be less frequent due to the reduced idiosyncrasy in the decision-making process, leading to more robust policies against shifting contingencies. Moreover, diverse boards tend to adopt more conventional financial policies, possibly including less risk, which reduce the dependence on firm debt and result in sustainably higher dividend yields for shareholders without harming the firm's organic growth (Bernile et al., 2018, p.602).

Furthermore, Bernile et al. (2018) argue that heterogeneous boards are more likely to invest in innovation projects that foster firm growth, even though R&D investments are typically riskier. This focus on innovation can be explained by the fact that such boards prefer more prudent risk-taking via the original concepts' promotion (Hoffman & Maier, 1961). There are also indications that board diversity positively influences the quantity and the quality of the firm's innovation output, measured by the number of patents or the ratio of patents to R&D expenses. Bearing in mind the board's advisory role, higher diversity could ex ante lead to a more efficient allocation of the firm's R&D resources. Specifically, diversity could promote more efficient monitoring of the firm's budget and resource allocation to more promising innovation areas (Bernile et al., 2018, p.603). Moreover, the *management theory* asserts that more diverse boards could positively shape corporate innovation practices through their impact on corporate culture (Griffin, Li, & Xu, 2021, p.127), as minority members of diverse boardrooms are more likely to challenge tradition, question the status quo, and inspire the majority members to adopt new perspectives (Johnson, van de Schoot, Delmar, & Crano, 2015, p.582). Lastly, board diversity generally fosters a diversity-friendly culture in the firm, thereby increasing the workforce's heterogeneity, which is essential for the firm's innovation process (Gao & Zhang, 2014).

Finally, Tarus and Aime (2014) examine the impact of board diversity on the firm's strategic change activities. Since the board is responsible for shaping the corporate strategic direction and reviewing progress in its implementation, the authors argue that different demographic and cognitive diversity characteristics might influence the firm's strategic change, defined as the change of the firm's resource allocation pattern. Next to the arguments outlined above, the authors add that younger people, having a fresher educational background, are more likely to expend more physical and mental effort on supporting the change and growth of their firms. Moreover, educational and functional diversity might help the boards spot environmental opportunities, and search and process comprehensive information more efficiently, translating them into viable strategies and ideas, and expanding the probability of accepting strategic change (Hambrick & Mason, 1984).

3. Descriptive Analysis of Diversity Trends in Germany

3.1. Data Description and Methodology

After the extensive literature review concerning the possible determinants and effects of board diversity, the focus moves to the central subject of this thesis, namely the board diversity trends in Germany. To analyse such trends in supervisory and executive boards, I use the BoardEx databank to raise data on all German listed firms from 1998 to 2020. To construct the different board diversity indices, I use information from two separate datasets: the first one (henceforth Main Set) includes the directors' general traits, such as their gender, birthday, corporate title, and the number of simultaneous board tenures, and the second one (henceforth Auxiliary Set) comprises information on their academic degrees, the award date, and their university affiliation.

To conduct the study, I use the STATA software, considering only observations from 1999 to 2019, as the years 1998 and 2020 consist of a small sample of firms, damaging cross-year comparability and precision. Furthermore, I drop companies with a foreign ISIN number to ensure that the trends focus only on firms that abide by the German corporate standards. Moreover, I exclude observations where the executive or supervisory boards consist of only one person, as such boards are homogenous per definition and can distort the trends.

Since diversity has various facets, and its dimensions may describe either demographic or cognitive differences among individuals (Williams & O'Reilly III, 1998), in my analysis, similar to Bernile et al. (2018), the main variables of interest are the gender, nationality, and age of board members, representing demographic attributes, as well as their financial expertise, the number of additional board tenures, and the institution, where each person received her latest academic degree, which provide information on the directors' cognitive characteristics.

While studying the diversification trends, it is essential to identify whether a specific diversity dimension has a more prominent driving force for the overall trend. Therefore, I proceed by studying three different levels of diversity. First, I

observe the overall diversity, which considers all six traits simultaneously for each board type (supervisory, executive or combined boards), following Bernile et al. (2018) in *Section 3.1.1*.[3] Then, I isolate the demographic and cognitive trends by constructing indices explained in *Section 3.1.2*, and lastly, I consider each dimension separately to identify their noteworthy trends. Depending on the level of aggregation, I employ different restrictions to exclude missing values. The analysis of the overall diversity comprises 15,129 observations for the supervisory boards, 6,100 for the executives, and 22,670 for the combined ones. *Appendix A* provides the total number of observations for each disaggregated study.

The variables for gender and nationality already exist in the dataset and are employed directly. To identify each board member's age, I build the difference between this person's birth year and the respective fiscal year for each tenure. To eliminate extreme outliers, distorting the trends, I winsorise the age variable at the 1% level. For the financial expertise dimension, I create a binary variable that takes the value one if a director is a financial expert. To do so, I consider each board member's role, looking for any keyword that could denote financial proficiency. For example, I identify a financial expert if they have a role description that includes terms like "CFO", "Risk", or the letter sequence "Fin".[4] Moreover, to study the number of other boards in which a board member sits simultaneously, I also construct a new variable. As the dataset contains the total number of boards on which each director serves each year, I subtract one board to consider only the *additional* board incumbencies.

Finally, for the educational diversity, I conduct several steps to prepare the variable of interest, which is the directors most recent university attendance. Following Bernile et al. (2018), I consider the institution where each member graduated as a proxy for education, but since many directors have attended multiple courses, I believe it is crucial to consider the latest university affiliation available in each reporting year. Specifically, I advocate that each university conveys a certain mentality to its students and that these experiences might influence the directors' mindset, work attitude, and extent of knowledge that they use in their job. Assuming that the most recently conveyed mentality probably has the most vivid effect on the director's perspectives, I use the latest university attendance as a proxy for educational diversity. To create this measure, firstly, I prepare the Auxiliary Set. Many board members have attained multiple degrees in the same year, and so I only keep the highest one. Moreover, in the Main Set, the oldest report year is 1999, and so if a director has several degrees attained before 1999, I keep the latest one. If a director, for example, has two degrees, one in 1970 and one in 1990, I only keep the one in 1990. If a director has a degree in 1990 and one in 2005, I keep both, and so, I undertake more steps to ensure that they are allocated correctly to the respective board years. The Auxiliary and the Main Set are joined together by each director's ID, and so, in each year, in each firm, each director receives all their degrees, meaning that duplicates can emerge. Thus, I drop all duplicates where the degree's award year is later than the respective report year, eliminating false joints. If all degrees are acquired before the report year, I only keep the latest one, according to my argumentation line above. For example, if a director has two degrees, one in 1990 and one in 2005, and the report year is 2000, only the 1990-degree is relevant. Yet, in 2010, both degrees are valid, but I only keep the one in 2005. Hence, each director receives the most recent university affiliation available in each year.

3.1.1. Board Diversity Index following Bernile et al. (2018)

To analyse the overall board diversity, similar to Bernile et al. (2018), I construct an analogous diversity index, considering all six dimensions. Following the authors, to compute the board diversity index, I calculate for each board-year observation the fraction of female directors (PCT_FEMALE), the mean number of other boards on which current members serve (NUM_BOARDS), the standard deviation of the directors' age (STDEV_AGE), and the Herfindahl–Hirschman Indices (henceforth HHI) of nationality (HHI_NATIONALITY), the most recent university affiliation (HHI_UNIVERSITY), and the binary variable for financial expertise (HHI_FINEXPERT). The authors favour the standard deviation of age over the HHI based on different age groups, as this approach does not cause mechanical changes in age diversity due to natural ageing.

Additionally, the authors use the HHI to compute the diversity level, as this is a standard method to measure the concentration within a specific group of observations, such as a board of directors. The standard HHI is defined as the sum of the squares of different group shares within the whole group, as showed in *Equation* 1:

$$HHI = \sum_{i=1}^{N} \left(s_i^2\right), \quad HHI \in \left[\frac{1}{N}; 1\right] \tag{1}$$

where s_i is the share of each category i, and N is the number of categories within a specific dimension. The HHI's value is limited between $\frac{1}{N}$ and one, where one indicates group homogeneity and $\frac{1}{N}$ perfect heterogeneity (Fahrmeir, Heumann, Künstler, Pigeot, & Tutz, 2016, pp. 79-80).[5]

Finally, the authors standardise each diversity attribute over the entire timespan to make their scales comparable and observe whether the diversity is above or below the average value of the 21-year-period. As Bernile et al. (2018) argue, each diversity component has equal importance for the final BOARD_DIVERSITY_INDEX in each board-year, presented in

[3] The term *combined boards* refers to the consideration of the supervisory and executive boards as if they were one boardroom, subject to the same sample restrictions. Thus, the results for the combined boards presented in the following sections do not necessarily average the diversity trends of the supervisory and executive boards, as the indices are computed anew over a larger board size.

[4] All relevant keywords are in *Appendix B*.

[5] Examples for the calculation of the HHI-based measures, as well as an extensive overview of all diversity measures can be found in *Appendix B*.

Equation (2), which is why I ensure that there are no missing values for any of the relevant variables when building the index.

$$\begin{aligned} BOARD\_DIVERSITY\_INDEX \\ &= STDZ(PCT\_FEMALE) \\ &+ STDZ(STDEV\_AGE) \\ &+ STDZ(NUM\_BOARDS) \\ &- STDZ(HHI\_NATIONALITY) \\ &- STDZ(HHI\_UNIVERSITY) \\ &- STDZ(HHI\_FINEXPERT) \end{aligned} \tag{2}$$

The authors propound to subtract HHI-based measures because, per definition, higher values reflect a higher concentration of the corresponding attribute in the board and, so, lower diversity. Thus, the higher the value of the board diversity index, the higher the diversity in the board.

3.1.2. Description of Other Variables

For the separate cognitive and demographic diversity trends' analysis, I follow the same steps, distinguishing between demographic attributes (gender, age, nationality) and cognitive traits (board tenure, financial expertise, most recent university affiliation). Thus, for each board-year, I observe either the fraction of female board members, the variance in age, and the HHI of nationality, or the mean number of additional boards and the HHIs of university and financial proficiency, as shown in *Equations* (3) and (4):

$$\begin{aligned} BOARD\_DEMOGRAPHIC\_DIVERSITY\_INDEX \\ = STDZ(PCT\_FEMALE) + STDZ(STDEV\_AGE) \\ - STDZ(HHI\_NATIONALITY) \end{aligned} \tag{3}$$

$$\begin{aligned} BOARD\_COGNITIVE\_DIVERSITY\_INDEX \\ = STDZ(NUM\_BOARDS) - STDZ(HHI\_UNIVERSITY) \\ - STDZ(HHI\_FINEXPERT) \end{aligned} \tag{4}$$

Lastly, it is also vital to analyse the trend of each attribute separately to identify whether any dimension is a more prominent diversity driver. For this final study, I conduct the same data preparation as before; however, I observe one trait at a time, eliminating missing values only for this dimension. In addition, I do not standardise the values, as I do not have to compare or combine them. Lastly, for the components where HHI was previously used, I employ the Blau's Index to make the diversity development more illustrative. The only difference between the two indices is that the Blau's Index is a transformation of the HHI, namely $1 - HHI$, and so higher values indicate greater diversity (Blau, 1977). The corresponding equation for each disaggregated diversity dimension in each board-year is shown below. The percentage of female board members, average number of board tenures, and the standard deviation of age remain the same.

$$BLAUS\_INDEX\_NATIONALITY = 1 - \sum_{n=1}^{N} s_n^2 \tag{5}$$

$$BLAUS\_INDEX\_UNIVERSITY = 1 - \sum_{u=1}^{N} s_u^2 \tag{6}$$

$$BLAUS\_INDEX\_FINEXPERT = 1 - \sum_{f=1}^{N} s_f^2 \tag{7}$$

3.2. Diversity Trends in Germany

3.2.1. Board Diversity Trends

To exhibit the trends, I present the computed values of the overall BOARD_DIVERSITY_INDEX in *Table* and *Figure* 1, averaging each index over all boards for each year. Studying German executive and supervisory boards combined, many companies tried to promote diversity in their boards from 1999 until 2002. Nevertheless, from 2003 until 2011, the diversity index gradually decreased, meaning that boards were becoming more homogeneous. Especially, from 2004 to 2012, the index is negative, indicating that the level of diversity was below average when considering the 21-year-span. Then, however, recognising the boards' diversification importance, companies strove to increase diversity until 2013. The years 2014 and 2015 present general negative board diversification trends, followed by an increase between 2015 and 2016. The years 2017 and 2018 were highlighted by a relative decrease in the heterogeneity of the directors' traits, which was, however, short-lived, as an increasing diversity trend can be observed since 2018.

When separating German supervisory and executive boards, similar conclusions emerge. The German supervisory boards' analysis mainly indicates a slight increase in the overall diversity in the first three years of the observed period. Next, a continuous decreasing trend of diversity can be identified until 2010. After this, the shareholders probably started attending who represents their interests and monitors business processes in firms, which also influenced the increase in the overall supervisory board diversity. However, until 2012, the index was continuously negative, indicating that the diversity levels were below the 21-year-average. Between 2013 and 2014, there was anew a sharp decrease in the board diversification, but an even more notable increase followed this until 2019.

As for the German executive boards, from 1999 until 2001, they experienced an increase in their overall heterogeneity. However, from 2002 until 2010, the diversity trend in the boardrooms was mainly downward, with sporadic short-term increases. Nevertheless, similarly to the diversity trend in the supervisory boards, the year 2010 was the board diversification strategy's turning point, marking a progressive increase until 2013. Afterwards, a short-term board heterogeneity decrease followed until 2015, but since then, board diversity has been rising uninterruptedly, also being above the 21-year-average since 2016.

Table 1: Board Diversity Index

Combined Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	-0.50	-0.30	0.01	0.28	0.05	-0.07	-0.09	-0.08	-0.10	-0.19	-0.06
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.03	-0.20	-0.02	0.07	-0.10	-0.15	0.14	0.09	0.11	0.56	
Supervisory Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	-0.36	-0.29	0.02	-0.05	-0.12	-0.13	-0.23	-0.26	-0.12	-0.45	-0.45
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.55	-0.30	-0.16	0.04	-0.04	0.06	0.29	0.41	0.45	0.71	
Executive Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	-0.23	0.22	0.39	0.34	0.09	0.24	0.06	-0.13	-0.16	-0.44	-0.33
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.47	-0.32	-0.08	0.25	0.14	-0.28	0.02	0.05	0.30	0.43	

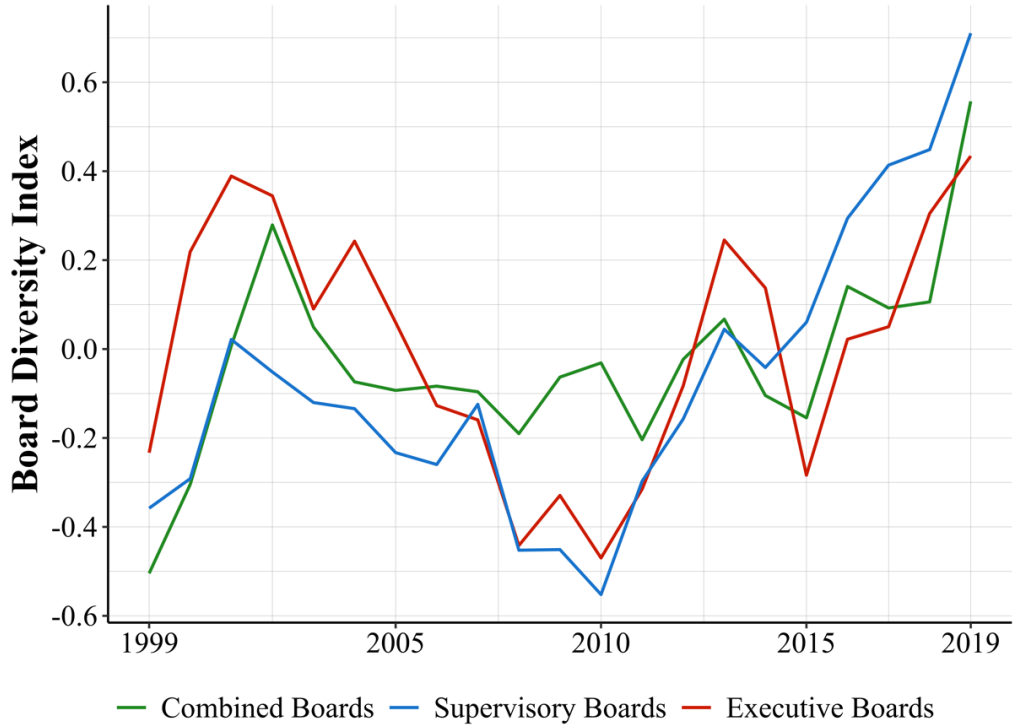


Figure 1: Board Diversity Index

Notes. The table and the figure present the development of board diversity in German listed companies. The Board Diversity Index is computed using data on gender, age, nationality, financial expertise, university affiliation, and the number of other board tenures, following Bernile et al. (2018). All components are standardised over the entire period so that the computed Board Diversity Index could provide information regarding the yearly diversity level relative to other years. Thus, the negative values mean that the diversity in this particular year is below the 21-year-average. All values in the table are rounded to two decimals, but the graph is plotted with higher decimal-precision.

3.2.2. Demographic and Cognitive Diversity Trends

After examining the overall board diversity trends, it is interesting to see what drives these tendencies and whether there are any visible differences between the developments of the demographic and cognitive diversity dimensions. *Table* and *Figure* 2 present these trends for German firms at the separate level.

Starting with the cognitive diversity, it is evident that the trend could be characterised by a general decrease, and since 2010, the level of cognitive diversity has remained under its 21-year-mean. The firms tended to have comparably heterogeneous boards at the beginning of the 21st century; however, after 2002, the combined boards started evincing more homogeneity regarding their cognitive traits. Despite some short-term fluctuations between 2011 and 2016, the negative trend has persisted.

The decreasing cognitive diversity trend can be identified for German supervisory boards, too, as board members have

Table 2: Cognitive and Demographic Board Diversity Indices

Cognitive Board Diversity Index											
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Combined Boards	0.83	0.97	0.96	0.97	0.84	0.73	0.47	0.33	0.14	0.04	0.03
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.05	-0.27	-0.25	-0.19	-0.18	-0.32	-0.26	-0.32	-0.37	-0.25	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Supervisory Boards	1.08	1.20	1.07	0.87	0.75	0.79	0.48	0.48	0.26	0.15	0.03
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.10	-0.37	-0.38	-0.22	-0.17	-0.23	-0.27	-0.35	-0.41	-0.31	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Executive Boards	0.46	0.75	0.74	0.66	0.41	0.50	0.12	0.04	0.00	-0.16	-0.02
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.14	-0.17	-0.11	-0.03	-0.15	-0.26	-0.13	-0.19	-0.28	-0.15	
Demographic Board Diversity Index											
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Combined Boards	-0.76	-0.69	-0.65	-0.57	-0.60	-0.63	-0.54	-0.43	-0.36	-0.21	-0.15
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.16	-0.06	0.20	0.29	0.17	0.19	0.40	0.45	0.53	0.59	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Supervisory Boards	-0.74	-0.68	-0.73	-0.73	-0.81	-0.79	-0.74	-0.68	-0.51	-0.41	-0.32
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.33	0.01	0.31	0.38	0.25	0.36	0.53	0.67	0.74	0.77	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Executive Boards	-0.21	-0.07	-0.08	-0.01	-0.05	-0.03	0.01	-0.11	-0.12	-0.06	-0.03
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	-0.05	-0.15	-0.07	0.00	0.00	-0.14	0.12	0.24	0.36	0.28	

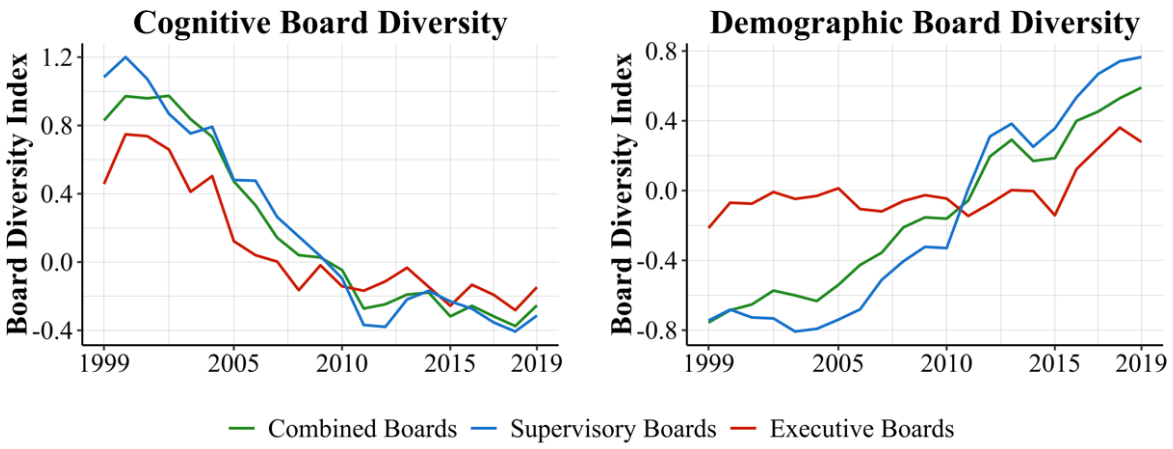


Figure 2: Cognitive and Demographic Board Diversity Indices

Notes. The tables and the figures present the development of the disaggregated cognitive and demographic board diversity characteristics in German listed companies. The Cognitive Board Diversity Index is computed using data on financial expertise, university affiliation, and the number of other board tenures; the Demographic Board Diversity Index is computed using the information regarding gender, age, and nationality of the board members, following Bernile et al. (2018). All components are standardised over the entire period so that both Board Diversity Indices could provide information regarding the yearly diversity level relative to other years. Thus, the negative values mean that the diversity in this particular year is below the 21-year-average. All values in the tables are rounded to two decimals, but the graphs are plotted with higher decimal-precision. The number of observations for each board diversity study is presented in *Appendix A*.

become less diverse in terms of their cognitive characteristics over the years. Between 2001 and 2012, this decline was also almost monotonous. Moreover, from 2012 to 2014, there were some attempts to increase board diversity, but these were probably not successful as the declining trend appeared again, lasting until 2018.

A similar trend can also be identified for the executive boardrooms. Diversity among executives was falling sharply until 2008, and this negative trend continued afterwards as well, albeit with a flatter slope. Since 2007, the level of diversity has also been constantly below the 21-year-average.

Switching to the demographic attributes, a definite positive trend can be noticed over the last 20 years for the combined boards, as they tended to become more heterogeneous regarding their members' gender, age, and nationality, and since 2012 the level of demographic board diversity is over its 21-year-average. Nevertheless, three time periods can be described as short-term exceptions, as the years 2002 to 2004, 2009 to 2010, and 2013 to 2014 represent minor decreases in the boardrooms' composition dissimilarity.

Similar positive trends can also be identified when studying German supervisory boards. Even though at the turn of the millennium, demographic board diversity was falling in such boards, an upward trend prevailed in 2003, and since then, only two modest declines have been noted in the years 2009 to 2010 and 2013 to 2014.

Lastly, as regards executive boards, the demographic diversity trend is not that prominent. These boards experienced systematic fluctuations until 2015, and thus, no clear trend can be identified. However, since 2015, there has been an increase in demographic heterogeneity, and the diversity level has been steadily above average since 2016.

3.2.3. Trends in Individual Diversity Categories

After discovering that it is rather demographic diversity that drives the overall diversity trends in German boardrooms, it is essential to learn the tendencies of each component of the BOARD_DIVERSITY_INDEX over time. Gender board diversification, being probably the most studied characteristic in the literature, could be seen as one of the most prominent examples of an upward trend, as presented in *Table* and *Figure* 3. At the combined boards level, one can observe a gradual positive trend until 2010, and then a sharp increase in female board representation. Similarly, studying the supervisory boards, the trend has had an overall rising character since 2002 and after 2015, there is an even more evident positive trend, which can be explained with the Gender Quota Act enacted by the government in 2016. In the executive boards, the trend is, however, not that explicit. While the overall tendency of the women fraction in the boards has a positive inclination, this development is rather S-shaped, with fluctuation ranging between almost identical maximum and minimum values.

A similar overall positive trend can be observed for nationality diversity. With some short-term exceptions, the combined, executive, and supervisory boards have been increasing hiring people from abroad, probably trying to capture the diversity benefits, discussed in *Section 2.5*.

The age diversity trends presented in *Table* and *Figure* 4 provide evidence that the combined boardrooms are becoming more age homogeneous over time. After 2006, when a weak positive trend is observed, the board's age heterogeneity has started declining, reaching its minimum in 2019. Similarly, the supervisory boards also present a negative trend regarding age diversity. The period between 2003 and 2008 can be described as the only relatively long-term positive trend in age deviations, followed, however, by a noticeable declining tendency until 2019. For executive boards, the decline had already started in 2002. Additionally, in the executive boardrooms the standard deviation of age is also smaller than in the other boards, meaning that the overall age structure of the supervisory boards is more heterogeneous than this in the executive boardrooms.

As for the separate analysis of the cognitive diversity attributes, one can observe that the individual's university affiliation has a negative diversity trend over the observed period in all board types, as presented in *Table* and *Figure* 5. Nevertheless, it is worth noting that the education diversity values have remained very high in supervisory boards despite the decreasing character. Interestingly, board diversity regarding financial expertise, which is also presented in *Figure* 5, has a positive trend at the combined, supervisory, and executive board levels over the entire period.

Finally, a clear negative trend can also be identified when studying the diversity regarding the average directors' board incumbencies in Table and Figure 6. While the average number of additional boards in which a board member serves was close to one, during the last decade, this number decreased significantly afterwards, indicating that nowadays, most board members do not serve on multiple boards at the same time. This, in turn, can be understood as a decrease in the board experience, an aspect that may be critical while accomplishing the board duties, but also as a possible reduction of the groupthink effect, as discussed in *Section 2.5*.

4. Discussion of the Results

Studying the trends, one can discern that while some diversity dimensions demonstrate clear upward trends in German corporations, some other dimensions have become even more homogeneous over time. One must note that the German corporate governance regulation regarding the gender quota has undoubtedly affected the diversity trends since 2016, as illustrated in *Section 3.2*. As already discussed, one diversity attribute might impact the heterogeneity of other dimensions, which could be seen in the trends of the nationality and financial expertise dimensions, as I believe that they behaved similar to the gender diversity trends over the last years. Nevertheless, it is also prominent that the age and education diversity among board members has negative trends, meaning that regardless of gender, nationality or functional background, the boards tend to become homogeneous in these two dimensions. This, however, might affect the whole

Table 3: Gender and Nationality Board Diversity

Gender Board Diversity											
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Combined Boards	0.07	0.07	0.06	0.07	0.07	0.08	0.07	0.06	0.06	0.06	0.06
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.06	0.07	0.10	0.10	0.11	0.13	0.14	0.15	0.15	0.16	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Supervisory Boards	0.09	0.08	0.08	0.08	0.09	0.09	0.09	0.08	0.08	0.08	0.08
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.08	0.09	0.11	0.12	0.13	0.15	0.18	0.18	0.19	0.20	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Executive Boards	0.02	0.02	0.02	0.02	0.02	0.03	0.02	0.02	0.02	0.02	0.02
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.02	0.02	0.05	0.05	0.05	0.05	0.06	0.05	0.06	0.06	
Nationality Board Diversity											
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Combined Boards	0.13	0.13	0.14	0.14	0.13	0.13	0.14	0.16	0.18	0.20	0.21
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.22	0.22	0.24	0.24	0.23	0.24	0.24	0.25	0.26	0.25	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Supervisory Boards	0.11	0.11	0.12	0.12	0.11	0.11	0.12	0.13	0.17	0.18	0.19
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.19	0.20	0.23	0.22	0.21	0.22	0.22	0.23	0.24	0.24	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Executive Boards	0.12	0.13	0.14	0.13	0.14	0.14	0.13	0.15	0.17	0.18	0.18
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.19	0.17	0.19	0.19	0.19	0.19	0.19	0.20	0.22	0.22	

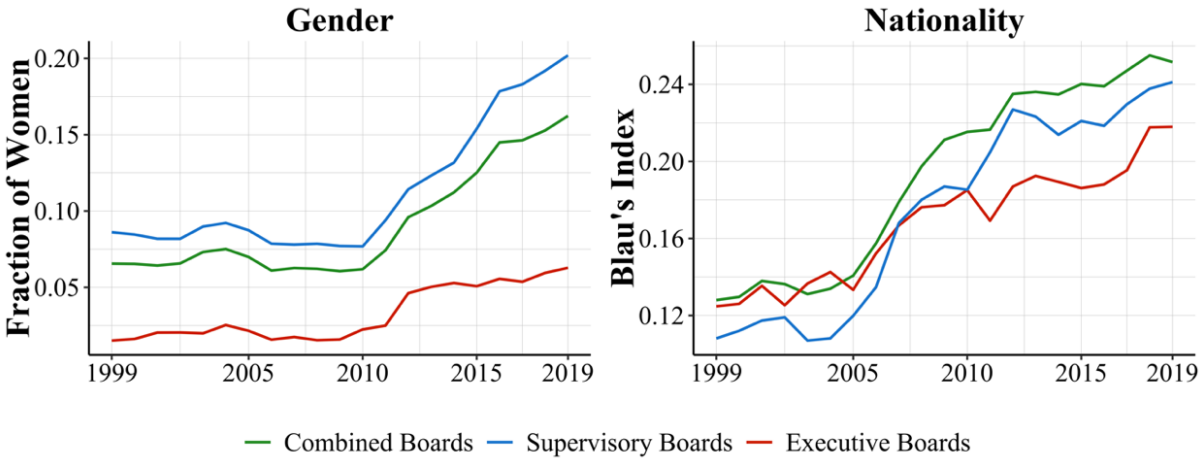


Figure 3: Gender and Nationality Board Diversity

Notes. The tables and the figures present the development of the separate gender and nationality diversity characteristics in German listed companies. The Gender Board Diversity represents the fraction of female board members relative to the number of all board members in each observed firm; the Nationality Diversity Blau's Index is computed using the information regarding the board members' nationality (Blau, 1977). All values in the tables are rounded to two decimals, but the graphs are plotted with higher decimal-precision. The number of observations for each board diversity study is presented in *Appendix A*.

boards' function and decision-making process, as many ideas and board rulings could be examined from a limited number of perspectives. Along the same argumentation lines as in *Section 2.5*, since differences in the board members' age and educational background might have positive impacts on firm performance and especially on risk appetite and risk exami-

Table 4: Standard Deviation of Board Members' Age

Combined Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	8.45	8.60	8.46	8.70	8.74	8.65	8.72	8.92	8.81	8.77	8.63
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	8.54	8.57	8.44	8.24	8.20	8.16	8.16	8.01	7.97	7.95	
Supervisory Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	8.26	8.27	8.25	8.28	8.19	8.30	8.62	8.60	8.60	8.73	8.55
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	8.52	8.35	8.42	8.14	8.27	8.34	8.31	8.25	8.02	7.94	
Executive Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	6.26	6.62	6.36	6.58	6.34	6.18	6.34	5.77	5.56	5.49	5.62
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	5.39	5.24	5.16	5.17	5.15	4.91	5.05	4.91	5.17	5.12	

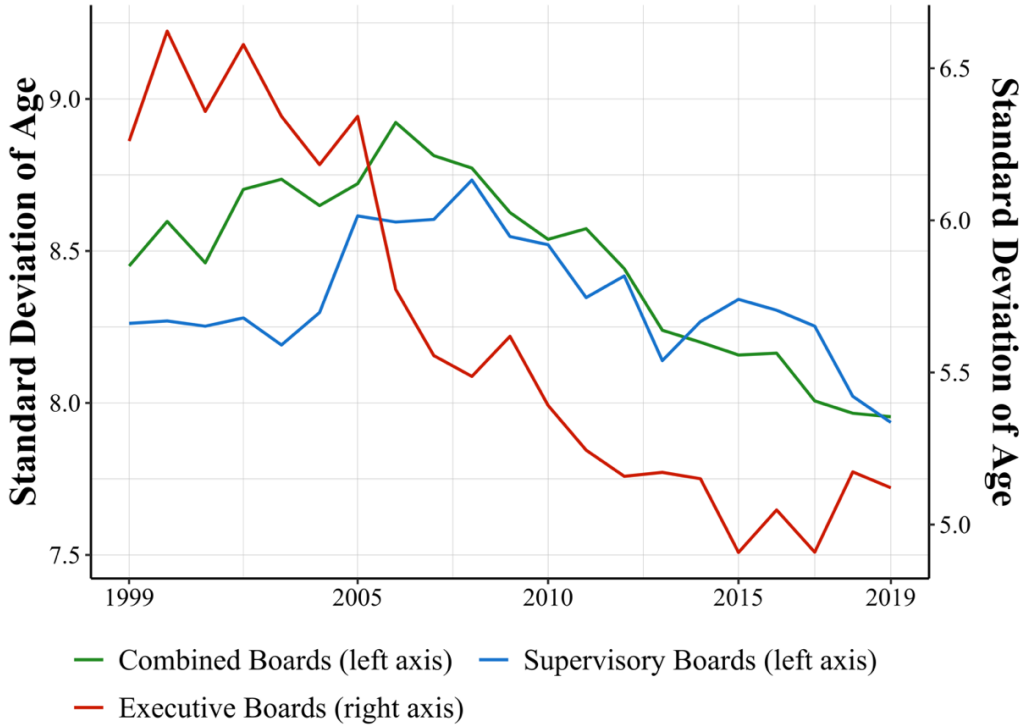


Figure 4: Standard Deviation of Board Members' Age

Notes. The table and the figure present the development of the distinct age diversity characteristic in German listed companies. The Standard Deviation of Board Members' Age represents the board-year average of the age standard deviations of its different members. All values in the table are rounded to two decimals, but the graph is plotted with higher decimal-precision. The number of observations for each age diversity study is presented in *Appendix A*.

nation methods, in light of these negative trends, one might consider an adverse effect on particular measures of company success. However, after breaking down the overall diversity index into its components, Bernile et al. (2018, p.590) expose that no single element of diversity alone drives the relationship between board diversity and firm risk. Besides, the effect of the board diversity index on risk remains significantly negative when the authors combine all components. This is also confirmed by other researchers in this topic (Baranchuk & Dybvig, 2009), asserting that the overall decision-making process depends on the joined effect of different diversity dimensions and not on its distinct parts.

This perspective could also explain the disagreement of previous empirical results that attempted to analyse the effects of specific board diversity attributes on firm performance, innovation processes, et cetera. Nevertheless, further research on board diversity is essential, as numerous determinants might influence its outcome. Thus, the firm's external environment, each country's specific economic and legislative setting, and the influence of globalisation could prompt

Table 5: Financial Expertise and University Affiliation Board Diversity

Financial Expertise Board Diversity											
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Combined Boards	0.06	0.08	0.08	0.09	0.11	0.13	0.13	0.15	0.15	0.16	0.16
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.16	0.18	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Supervisory Boards	0.00	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.04	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.07	0.07	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Executive Boards	0.14	0.18	0.20	0.22	0.26	0.28	0.29	0.30	0.30	0.31	0.32
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.32	0.33	0.35	0.34	0.34	0.35	0.35	0.34	0.34	0.34	
University Affiliation Board Diversity											
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Combined Boards	0.77	0.77	0.79	0.80	0.80	0.79	0.77	0.77	0.76	0.76	0.77
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.77	0.74	0.74	0.75	0.76	0.75	0.76	0.76	0.75	0.77	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Supervisory Boards	0.71	0.72	0.72	0.72	0.74	0.73	0.72	0.74	0.72	0.72	0.72
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.72	0.69	0.69	0.71	0.71	0.71	0.71	0.71	0.70	0.71	
	1999	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
Executive Boards	0.68	0.68	0.68	0.67	0.66	0.66	0.63	0.64	0.62	0.62	0.63
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.61	0.60	0.62	0.63	0.61	0.60	0.61	0.60	0.59	0.60	

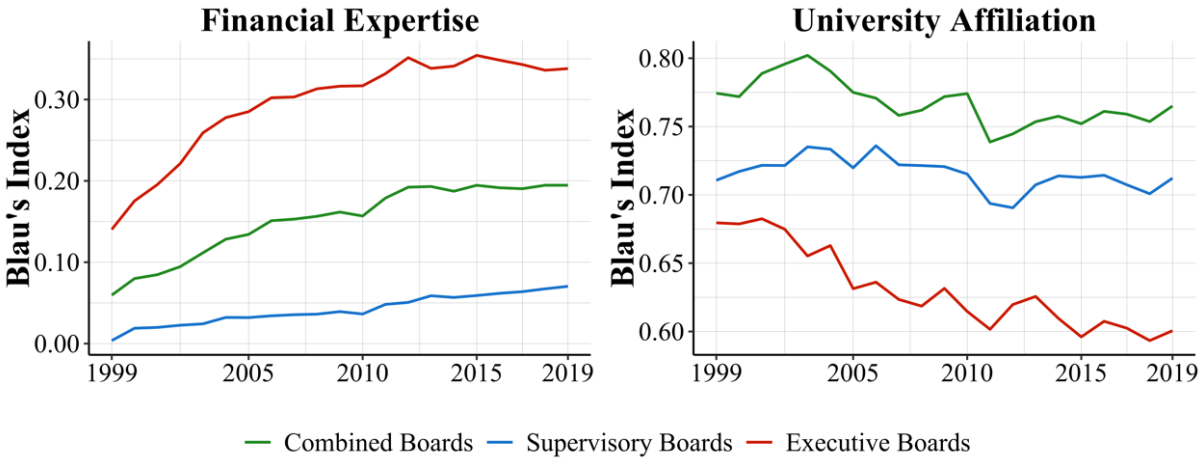


Figure 5: Financial Expertise and University Affiliation Board Diversity

Notes. The tables and the figures present the development of the distinct financial expertise and university affiliation diversity characteristics in German listed companies. The Financial Expertise Board Diversity represents the board-year average values of the Blau's Index regarding the number of financial experts in each board; the University Affiliation Blau's Diversity Index is computed using the information regarding the board members' visited universities, in which they received their latest academic degree (Blau, 1977). All values in the tables are rounded to two decimals, but the graphs are plotted with higher decimal-precision. The number of observations for each board diversity study is presented in *Appendix A*.

diversity trends with its benefits and costs. That is why I would like to give rise to possible questions and propositions for future research, as the issue of board diversity might become even more critical in the following years because of its ethical and economic reasoning. As the focus of this thesis is the diversity trends in *German* listed firms, I would also

Table 6: Mean number of boards in which the directors sit simultaneously

Combined Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	0.89	0.91	0.89	0.84	0.75	0.75	0.61	0.56	0.52	0.48	0.44
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.43	0.38	0.38	0.38	0.36	0.34	0.34	0.34	0.33	0.33	
Supervisory Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	1.07	1.07	1.04	0.98	0.90	0.89	0.73	0.68	0.64	0.59	0.55
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.53	0.47	0.48	0.47	0.46	0.44	0.43	0.42	0.41	0.40	
Executive Boards	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
	0.53	0.60	0.56	0.53	0.47	0.46	0.39	0.33	0.31	0.28	0.26
	2010	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	
	0.24	0.21	0.21	0.21	0.20	0.18	0.17	0.19	0.18	0.19	

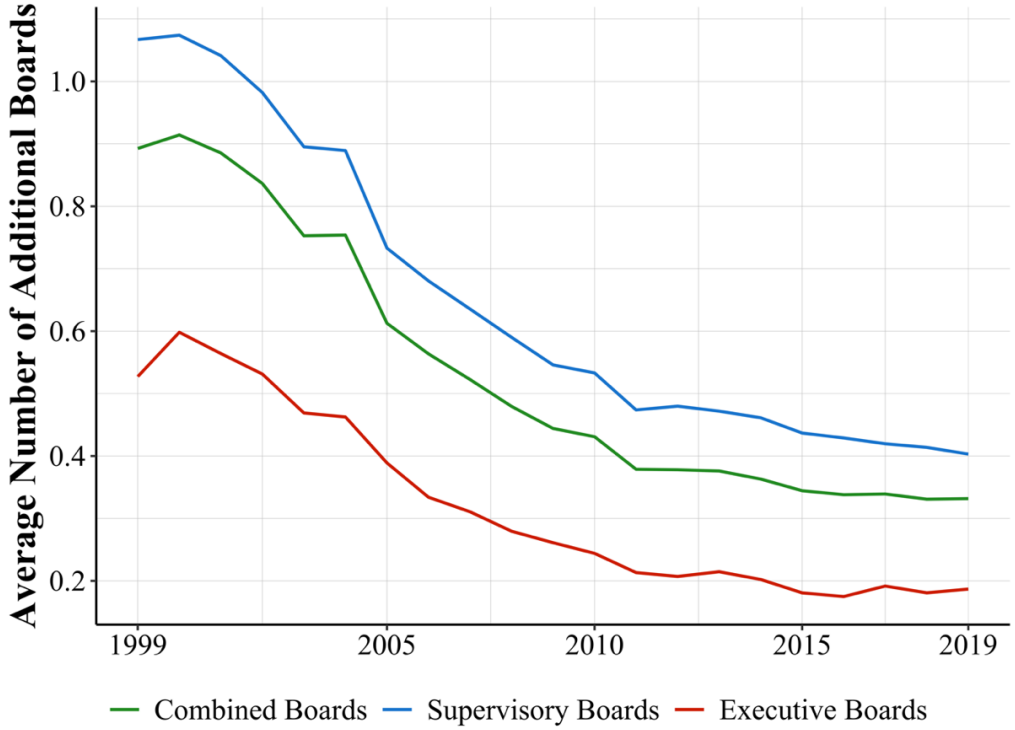


Figure 6: Mean number of boards in which the directors sit simultaneously

Notes. The table and the figure present the development of the distinct board incumbencies diversity characteristic in German listed companies. This diversity study represents the board-year average of the mean number of different boards, in which every board member serves simultaneously. All values in the table are rounded to two decimals, but the graph is plotted with higher decimal-precision. The number of observations for each board incumbencies diversity study is presented in *Appendix A*.

formulate my propositions considering this national context.

Firstly, as board diversity might define the degree to which idiosyncrasies in the board members' motives and the access to information influence the company-wide decision-making process, it should have a first-order impact on the corporate risk that stems from these decisions (Yousaf, Jebran, & Wang, 2021). Thus, board diversity could moderate arrangements and build a synthesis of multiple opinions and knowledge that could benefit companies operating in more volatile environments (Bernile et al., 2018, p.595).

Proposition 1. *Ceteris paribus, a higher level of overall diversity in the boardrooms of German companies is associated with lower firm risk.*

Furthermore, while one might state that a decreased level of risk could curtail shareholder value, the diversity in experience, personal qualities, et cetera might lead to greater monitoring activities and more thorough risk and competitive strategies (Carter et al., 2003; Bernile et al., 2019). This, in turn, could result in advantages for firm profitability and

value, as formulated in the following proposition.

Proposition 2. *Ceteris paribus, there is a positive relationship between overall board diversity and firm performance in German companies.*

Additionally, the enhanced board monitoring activity might also result in a more proper design of the firm's incentive systems and the compensation packages for the executives (Adams & Ferreira, 2009). Moreover, the various backgrounds of the directors could also affect the amount of compensation, as due to globalisation, foreign board members may stimulate the level of salaries, making them globally comparable (Randøy & Nielsen, 2002).

Proposition 3. *Ceteris paribus, a higher level of overall diversity in the boardrooms of German companies is associated with better incentive systems and higher board compensations.*

Lastly, as it has been stated, diversity in backgrounds and experiences could be vital when boardrooms demand creative and novel solutions (Miller & del Carmen Triana, 2009, pp.759-761). As innovation is crucial for the success of many companies, studying the effect of board diversity on R&D investments, which are an essential part of the firm's innovation process, can shed light on the additional effects of diverse boards (Bernile et al., 2018).

Proposition 4. *Ceteris paribus, there is a positive relationship between overall board diversity and the quality and quantity of R&D expenditures in German companies.*

5. Conclusion

Corporate governance theory contends that the board structure strongly influences the board's actions that ultimately affect firm performance (Kim, Burns, & Prescott, 2009). The empirical literature has provided mixed results and conceptual arguments regarding board diversity and its effects on various firm outcomes. That is why Milliken and Martins (1996, p.403) call board diversity "a double-edged sword", as the enhanced creativity and the variety of skills and experiences could also backfire if board members become dissatisfied and fail to identify with the rest of the group, causing conflicts and group fragmentation (Wright & Snell, 1999, p.50). Bearing these aspects in mind, the focus of this thesis was to examine the potential effects of board diversity (and its different facets) on firm performance, firm risk, and other firm outcomes. It becomes apparent that while boardroom diversity might improve the board's monitoring ability, moderate the decision-making process, and foster innovation, the firm's external environment often has the final word when determining whether board diversity is beneficial or rather costly for a specific company. Along the same lines, many researchers denote the importance to boost the different dimensions of diversity simultaneously, emphasising their uniqueness and non-interchangeability, since some diversity components may create more powerful synergies when combined (Pelled, 1996; Williams & O'Reilly III, 1998).

Today, many companies in the developed world endeavour to promote diversity. Public and academic institutions do their best to urge diverse workforces and management teams to introduce, for instance, gender quotas for their corporate regulations (Eckbo, Nygaard, & Thorburn, 2020). That could explain the interest of this thesis to investigate whether there are any evident board diversity trends in Germany, being a country that values equality (Rohrschneider, 1999). After discussing firm internal and external factors that may drive board diversity, the empirical part follows, presenting the actual trends of the overall as well as separate diversity dimensions. One could note that while the general board diversity in German firms has risen since 1999, the main driving forces of this positive trend is the demographic variety between board members, especially regarding gender and nationality. Notwithstanding, some dimensions have been increasingly characterised by homogeneity, as board members tend to have degrees from the same universities or belong to similar age groups.

Albeit proposing exciting paths for future research, this study has some limitations. First, although the award year and the universities are recorded for most of the directors, the degree they obtain is in many cases not registered, prohibiting the analysis of the diversity regarding educational attainments. Thus, following Bernile et al. (2018), who proxy education via the academic institutions where the directors received their bachelor's degree, and adjusting this measure in order to overcome the limitation of the dataset, I consider the educational diversity in terms of the universities where the directors attained their latest degree. However, even with this circumvention, I believe that analysing the variation in the level of the directors' qualifications or even their fields of study could be very valuable to research. That way, the actual difference in the cognitive capacities and knowledge, which are crucial determinants of the monitoring and information processing capabilities (Mahadeo et al., 2012, p.378), could be sufficiently captured.

Moreover, the employed datasets also pose further limitations, as they contain many missing values for various characteristics and, in many cases, incomplete information. That is why in the study of the overall board diversity trend, the number of observations used is comparatively low. Contrarywise, the analyses of the separate diversity dimensions are conducted with much larger samples since fewer restrictions for missing values apply. Yet, this may impede the indices' comparability because of the differences in the population size and damage the precision of the overall index owning to the relatively smaller sample.

In addition, future studies could emphasise other diversity attributes, such as the directors' religion, native tongue, or political preferences. These individual traits might also influence in-board interactions and corporate effectiveness (Carter et al., 2010, p.411), and thus, the research could deliver insights on the optimal board constellation.

Finally, the study's time span covers the years between 1999 and 2019, and only the last three years of this period are affected by the German gender quota regulation. As discussed, the different drivers of diversity are interconnected and interdependent, and as such, it may be fascinating to analyse diversity trends for a more prolonged period after the mandatory quota introduction. Consequently, similar analyses could be conducted in the years to come, after the quota regulations have already rooted in the economy. This could deliver ground-breaking results regarding the long-term effects of board diversity, possibly giving new impulses for further discussions and subsequent policies, aiming to improve corporate governance processes and, consequently, our society.

References

Acharya, V. V. (2010). *Regulating wall street*. Hoboken, NJ: John Wiley and Sons.

Adams, R. B., de Haan, J., Terjesen, S., & van Ees, H. (2015). Board diversity: Moving the field forward. *Corporate Governance: An International Review*, *23*(2), 77–82.

Adams, R. B., & Ferreira, D. (2007). A theory of friendly boards. *The Journal of Finance*, *62*(1), 217–250.

Adams, R. B., & Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics*, *94*(2), 291–309.

Adams, R. B., & Funk, P. (2012). Beyond the Glass Ceiling: Does Gender Matter? *Management Science*, *58*(2), 219–235.

Adams, R. B., Hermalin, B. E., & Weisbach, M. S. (2010). The role of boards of directors in corporate governance: A conceptual framework and survey. *Journal of Economic Literature*, *48*(1), 58–107.

Ahern, K. R., & Dittmar, A. K. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation ∗. *The Quarterly Journal of Economics*, *127*(1), 137–197.

AktG. (1965). *German stock corporation act. bgbl. i.*

Ali, M., Kulik, C. T., & Metz, I. (2011). The gender diversity–performance relationship in services and manufacturing organizations. *The International Journal of Human Resource Management*, *22*(7), 1464–1485.

Anderson, R. C., Reeb, D. M., Upadhyay, A., & Zhao, W. (2011). The economics of director heterogeneity. *Financial Management*, *40*(1), 5–38.

Arena, C., Cirillo, A., Mussolino, D., Pulcinelli, I., Saggese, S., & Sarto, F. (2015). Women on board: evidence from a masculine industry. *Corporate Governance*, *15*(3), 339–356.

Arnegger, M., Hofmann, C., Pull, K., & Vetter, K. (2014). Firm size and board diversity. *Journal of Management & Governance*, *18*(4), 1109–1135.

Arrow, K. (1951). *Social choice and individual values*. New York: John Wiley & Sons.

Bantel, K. A. (1993a). Comprehensiveness of strategic planning: The importance of heterogeneity of a top team. *Psychological Reports*, *73*(1), 35–49.

Bantel, K. A. (1993b). Strategic clarity in banking: Role of top management-team demography. *Psychological Reports*, *73*(3_suppl), 1187–1201.

Baranchuk, N., & Dybvig, P. H. (2009). Consensus in diverse corporate boards. *Review of Financial Studies*, *22*(2), 715–747.

Bebchuk, L., & Fried, J. (2004). *Pay without performance: The unfulfilled promise of executive compensation*. Cambridge, MA: Harvard University Press.

Becker, G. S. (1976). *Human capital*. New York: National Bureau of Economic Research.

Benkraiem, R., Hamrouni, A., Lakhal, F., & Toumi, N. (2017). Board independence, gender diversity and CEO compensation. *Corporate Governance: The International Journal of Business in Society*, *17*(5), 845–860.

Berger, A. N., Kick, T., & Schaeck, K. (2014). Executive board composition and bank risk taking. *Journal of Corporate Finance*, *28*, 48–65.

Berglof, E. (2011). A european perspective on the global financial crisis. *Corporate Governance: An International Review*, *19*(5), 497–501.

Berle, A., & Means, G. (1932). *The modern corporation and private property*. New York: Harcourt, Brace & World.

Bernasek, A., & Shwiff, S. (2001). Gender, risk, and retirement. *Journal of Economic Issues*, *35*(2), 345–356.

Bernile, G., Bhagwat, V., & Yonker, S. (2018). Board diversity, firm risk, and corporate policies. *Journal of Financial Economics*, *127*(3), 588–612.

Bertrand, M., & Schoar, A. (2003). Managing with Style: The Effect of Managers on Firm Policies. *The Quarterly Journal of Economics*, *118*(4), 1169–1208.

BGBl. (2015). *Gesetz für die gleichberechtigte Teilhabe von Frauen und Männern an Führungspositionen in der Privatwirtschaft und im öffentlichen Dienst*. Teil I Nr. 17, Bundesgesetzblatt.

Bhat, K. U., Chen, Y., Jebran, K., & Memon, Z. A. (2019). Board diversity and corporate risk: evidence from China. *Corporate Governance: The International Journal of Business in Society*, *20*(2), 280–293.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure* (Vol. 7). New York: Free Press.

Bowles, S., & Gintis, H. (2002). Social capital and community governance. *The Economic Journal*, *112*(483), F419–F436.

Brammer, S., Millington, A., & Pavelin, S. (2007). Gender and ethnic diversity among UK corporate boards. *Corporate Governance: An International Review*, *15*(2), 393–403.

Byrne, D. (1971). *The attraction paradigm*. New York: Academic Press.

Cambridge Dictionary. (2021). *Meaning of human capital in english*. https://dictionary.cambridge.org/dictionary/english/human-capital. (Retrieved May 5, 2021)

Campbell, J. Y. (2006). Household finance. *The Journal of Finance*, *61*(4), 1553–1604.

Campbell, K., & Mínguez-Vera, A. (2008). Gender diversity in the boardroom and firm financial performance. *Journal of Business Ethics*, *83*(3), 435–451.

Carter, D. A., D'Souza, F., Simkins, B. J., & Simpson, W. G. (2010). The gender and ethnic diversity of US boards and board committees and firm financial performance. *Corporate Governance: An International Review*, *18*(5), 396–414.

Carter, D. A., Simkins, B. J., & Simpson, W. G. (2003). Corporate governance, board diversity, and firm value. *The Financial Review*, *38*(1), 33–53.

Casteuble, C., Lepetit, L., & Tran, T. T. (2019). *Women on boards: do quotas affect firm performance?* (Tech. Rep.). Working Papers hal-02385034, HAL, 1-45.

Chaganti, R. S., Mahajan, V., & Sharma, S. (1985). CORPORATE BOARD SIZE, COMPOSITION AND CORPORATE FAILURES IN RETAILING INDUSTRY[1]. *Journal of Management Studies*, *22*(4), 400–417.

Chen, L. H., Gramlich, J., & Houser, K. A. (2017). The effects of board gender diversity on a firm's risk strategies. *SSRN Electronic Journal*.

Coles, J. L., Daniel, N. D., & Naveen, L. (2014). Co-opted boards. *SSRN Electronic Journal*.

Daft, R. (2013). *Understanding the theory and design of organizations*. South-Western Cengage Learning.

Davidson, W. N., Xie, B., Xu, W., & Ning, Y. (2007). The influence of executive age, career horizon and incentives on pre-turnover earnings management. *Journal of Management & Governance*, *11*(1), 45–60.

de Cabo, R. M., Gimeno, R., & Escot, L. (2011). Disentangling discrimination on spanish boards of directors. *Corporate Governance: An International Review*, *19*(1), 77–95.

Delis, M. D., Gaganis, C., Hasan, I., & Pasiouras, F. (2017). The effect of board directors from countries with different genetic diversity levels on corporate performance. *Management Science*, *63*(1), 231–249.

Dittmann, I., Maug, E., & Schneider, C. (2010). Bankers on the boards of german firms: What they do, what they are worth, and why they are (still) there∗. *Review of Finance*, *14*(1), 35–71.

Dumas, T. L., Phillips, K. W., & Rothbard, N. P. (2013). Getting closer at the company party: Integration experiences, racial dissimilarity, and workplace relationships. *Organization Science*, *24*(5), 1377–1401.

Eckbo, B. E., Nygaard, K., & Thorburn, K. S. (2020). *Valuation effects of norway's board gender-quota law revisited* (Tech. Rep.). Tuck School of Business Working Paper, (2746786).

Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *The Academy of Management Review*, *14*(1), 57–74.

Erhardt, N. L., Werbel, J. D., & Shrader, C. B. (2003). Board of director diversity and firm financial performance. *Corporate governance: An international review*, *11*(2), 102–111.

Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., & Tutz, G. (2016). *Statistik: Der Weg zur Datenanalyse*. Berlin: Springer-Verlag.

Fama, E. F., & Jensen, M. C. (1983). Agency problems and residual claims. *The Journal of Law and Economics*, *26*(2), 327–349.

Farag, H., & Mallin, C. (2018). The influence of CEO demographic characteristics on corporate risk-taking: evidence from chinese IPOs. *The European Journal of Finance*, *24*(16), 1528–1551.

Farrell, K. A., & Hersch, P. L. (2005). Additions to corporate boards: the effect of gender. *Journal of Corporate Finance*, *11*(1-2), 85–106.

Fauver, L., & Fuerst, M. E. (2006). Does good corporate governance include employee representation? evidence from german corporate boards. *Journal of Financial Economics*, *82*(3), 673–710.

Fernández-Temprano, M. A., & Tejerina-Gaite, F. (2020). Types of director, board diversity and firm performance. *Corporate Governance: The International Journal of Business in Society*, *20*(2), 324–342.

Fiedler, F. E. (1967). *A theory of leadership effectiveness*. New York: McGraw-Hill.

Finkelstein, S., Hambrick, D., & Cannella, A. A. (1996). *Strategic leadership*. St. Paul: West Educational Publishing.

Galbraith, J. (1973). *Designing complex organizations*. Readlng, MA: Addlson- Wesle.

Gao, H., & Zhang, W. (2014). *Does workforce diversity pay? evidence from corporate innovation* (Tech. Rep.). Unpublished working paper. Nanyang Technological University.

Gordini, N., & Rancati, E. (2017). Gender diversity in the italian boardroom and firm financial performance. *Management Research Review*, *40*(1), 75–94.

Graham, J. R., & Harvey, C. R. (2001). The theory and practice of corporate finance: evidence from the field. *Journal of Financial Economics*, *60*(2-3), 187–243.

Griffin, D., Li, K., & Xu, T. (2021). Board gender diversity and corporate innovation: International evidence. *Journal of Financial and Quantitative Analysis*, *56*(1), 123–154.

Guest, P. M. (2019). Does Board Ethnic Diversity Impact Board Monitoring Outcomes? *British Journal of Management*, *30*(1), 53–74.

Hafsi, T., & Turgut, G. (2013). Boardroom diversity and its effect on social performance: Conceptualization and empirical evidence. *Journal of Business Ethics*, *112*(3), 463–479.

Halko, M.-L., Kaustia, M., & Alanko, E. (2012). The gender effect in risky asset holdings. *Journal of Economic Behavior & Organization*, *83*(1), 66–81.

Hambrick, D. C., Cho, T. S., & Chen, M.-J. (1996). The influence of top management team heterogeneity on firms' competitive moves. *Administrative Science Quarterly*, *41*(4), 659–684.

Hambrick, D. C., & Mason, P. A. (1984). Upper echelons: The organization as a reflection of its top managers. *The Academy of Management Review*, *9*(2), 193–206.

Harrigan, K. R. (1981). Numbers and positions of women elected to corporate boards. *Academy of Management Journal*, *24*, 619-625.

Harris, M., & Raviv, A. (2008). A Theory of Board Control and Size. *The Review of Financial Studies*, *21*(4), 1797–1832.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, *32*(4), 1199–1228.

Haslam, S. A., Ryan, M. K., Kulich, C., Trojanowski, G., & Atkins, C. (2010). Investing with prejudice: the relationship between women's presence on company boards and objective and subjective measures of company performance. *British Journal of Management*.

Hillman, A. J., Cannella, A. A., & Harris, I. C. (2002). Women and racial minorities in the boardroom: How do directors differ? *Journal of Management*, *28*(6), 747–763.

Hillman, A. J., Cannella, A. A., & Paetzold, R. L. (2000). The resource dependence role of corporate directors: Strategic adaptation of board composition in response to environmental change. *Journal of Management Studies*, *37*(2), 235–256.

Hillman, A. J., Keim, G. D., & Luce, R. A. (2001). Board composition and stakeholder performance: Do stakeholder directors make a difference? *Business & Society*, *40*(3), 295–314.

Hillman, A. J., Shropshire, C., & Cannella, A. A. (2007). Organizational predictors of women on corporate boards. *Academy of Management Journal*, *50*(4), 941–952.

Hoffman, L. R., & Maier, N. R. F. (1961). Quality and acceptance of problem solutions by members of homogeneous and heterogeneous groups. *The Journal of Abnormal and Social Psychology*, *62*(2), 401–407.

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of Management*, *33*(6), 987–1015.

Hyland, M. M., & Marcellino, P. A. (2002). Examining gender on corporate boards: a regional study. *Corporate Governance: The international journal of business in society*, *2*(4), 24–31.

Jackson, S. E., May, K. E., Whitney, K., Guzzo, R. A., & Salas, E. (1995). *Understanding the dynamics of diversity in decision-making teams. team effectiveness and decision making in organizations*. San Francisco: Jossey-Bass.

Jackson, S. E., Schuler, R. S., & Rivero, J. C. (1989). Organizational Characteristics As Predictors Of Personnel Practices. *Personnel Psychology*, *42*(4), 727–786.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. In *Economics social institutions* (pp. 163–231). Springer Netherlands.

Johnson, A. R., van de Schoot, R., Delmar, F., & Crano, W. D. (2015). Social influence interpretation of interpersonal processes and team performance over time using bayesian model selection. *Journal of Management*, *41*(2), 574–606.

Johnson, S. G., Schnatterly, K., & Hill, A. D. (2013). Board composition beyond independence. *Journal of Management*, *39*(1), 232–262.

Kandel, E., & Lazear, E. P. (1992). Peer Pressure and Partnerships. *Journal of Political Economy*, *100*(4), 801–817.

Kang, H., Cheng, M., & Gray, S. J. (2007). Corporate governance and board composition: diversity and independence of australian boards. *Corporate Governance: An International Review*, *15*(2), 194–207.

Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *The American Journal of Sociology*, *82*(5), 695–990.

Khan, I., Khan, I., & bin Saeed, B. (2019). Does board diversity affect quality of corporate social responsibility disclosure? evidence from pakistan. *Corporate Social Responsibility and Environmental Management*, *26*(6), 1371–1381.

Kilduff, M., Angelmar, R., & Mehra, A. (2000). Top management-team diversity and firm performance: Examining the role of cognitions. *Organization Science*, *11*(1), 21–34.

Kim, B., Burns, M. L., & Prescott, J. E. (2009). The strategic role of the board: The impact of board structure on top management team strategic action capability. *Corporate Governance: An International Review*, *17*(6), 728–743.

Kim, H., & Lim, C. (2010). Diversity, outside directors and firm valuation: Korean evidence. *Journal of Business Research*, *63*(3), 284–291.

Klein, A. (2002). Audit committee, board of director characteristics, and earnings management. *Journal of Accounting and Economics*, *33*(3), 375–400.

Lane, C. (2003). Changes in Corporate Governance of German Corporations: Convergence to the Anglo-American Model? *Competition & Change*, *7*(2-3), 79–100.

Lawrence, P. R., & Lorsch, J. W. (1967). Differentiation and integration in complex organizations. *Administrative Science Quarterly*, *12*(1), 1–47.

Liao, L., Luo, L., & Tang, Q. (2015). Gender diversity, board independence, environmental committee and greenhouse gas disclosure. *The British Accounting Review*, *47*(4), 409–424.

Lucas-Pérez, M. E., Mínguez-Vera, A., Baixauli-Soler, J. S., Martín-Ugedo, J. F., & Sánchez-Marín, G. (2015). Women on the board and managers' pay: Evidence from spain. *Journal of Business Ethics*, *129*(2), 265–280.

Luoma, P., & Goodstein, J. (1999). Research Notes. Stakeholders And Corporate Boards: Institutional Influences On Board Composition And Structure. *Academy of Management Journal*, *42*(5), 553–563.

Luthy, D. H. (1998). Intellectual capital and its measurement. In *Proceedings of the asian pacific interdisciplinary research in accounting conference (apira)*. Osaka, Japan.

Mahadeo, J. D., Soobaroyen, T., & Hanuman, V. O. (2012). Board composition and financial performance: Uncovering the effects of diversity in an emerging economy. *Journal of Business Ethics*, *105*(3), 375–388.

McCauley, C. (1989). The nature of social influence in groupthink: Compliance and internalization. *Journal of Personality and Social Psychology*, *57*(2), 250–260.

Miller, T., & del Carmen Triana, M. (2009). Demographic diversity in the boardroom: Mediators of the board diversity-firm performance relationship. *Journal of Management Studies*, *46*(5), 755–786.

Milliken, F. J., & Martins, L. L. (1996). Searching for common threads: Understanding the multiple effects of diversity in organizational groups. *The Academy of Management Review*, *21*(2), 402–433.

Minton, B. A., Taillard, J. P., & Williamson, R. (2014). Financial Expertise of the Board, Risk Taking, and Performance: Evidence from Bank Holding Companies. *The Journal of Financial and Quantitative Analysis*, *49*(2), 351–380.

Monks, R. A., & Minow, N. (2004). Corporate governance. In 3rd (Ed.), . Madden, MA: Blackwell Publishing Ltd.

Nielsen, B. B., & Nielsen, S. (2013). Top management team nationality diversity and firm performance: A multilevel study. *Strategic Management Journal*, *34*(3), 373–382.

Ntim, C. G. (2015). Board diversity and organizational valuation: unravelling the effects of ethnicity and gender. *Journal of Management & Governance*, *19*(1), 167–195.

OECD. (2005). *Definition: Corporate governance. oecd: Glossary of statistical terms: https://stats.oecd.org/glossary/detail.asp?id=6778.* (Retrieved May 5, 2021)

Oxelheim, L., & Randøy, T. (2005). The Anglo-American Financial Influence on CEO Compensation in Non-Anglo-American Firms. *Journal of International Business Studies*, *36*(4), 470–483.

Oxelheim, L., Stonehill, A., Randøy, T., Vikkula, K., Dullum, K., & Modén, K. (1998). *Corporate strategies to internationalise the cost of capital*. Copenhagen: Copenhagen Business School Press.

Oxelheim, L., & Wihlborg, C. (1997). *Managing in the turbulent world economy: Corporate performance and risk exposure*. London and New York: Wiley.

Pelled, L. H. (1996). Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organization Science*, *7*(6), 615–631.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict and performance. *Administrative Science Quarterly*, *44*(1), 1–28.

Peltomäki, J., Sihvonen, J., Swidler, S., & Vähämaa, S. (2020). Age, gender, and risk-taking: Evidence from the S&P 1500 executives and market-based measures of firm risk. *Journal of Business Finance & Accounting*, *48*(9-10), 1988–2014.

Pfeffer, J., & Salancik, G. R. (1978). *The external control of organizations: A resource dependence perspective*. New York: Harper & Row.

Pfeffer, J., & Salancik, G. R. (2003). *The external control of organizations: A resource dependence perspective*. Stanford, CA: Stanford University Press.

Plaut, V. C. (2010). Diversity science: Why and how difference makes a difference. *Psychological Inquiry*, *21*(2), 77–99.

Randøy, T., & Nielsen, J. (2002). Company performance, corporate governance, and ceo compensation in norway and sweden. *Journal of Management and Governance*, *6*(1), 57–81.

Regierungskommission. (2019). *German corporate governance code.* https://www.dcgk.de//files/dcgk/usercontent/en/download/code/191216_German_Corporate_Governance_Code.pdf. (Retrieved May 5, 2021)

Rohrschneider, R. (1999). *Learning democracy: Democratic and economic values in unified germany*. Oxford: Oxford University Press.

Rose, C. (2007). Does female board representation influence firm performance? the danish evidence. *Corporate Governance: An International Review*, *15*(2), 404–413.

Rose, C., Munch-Madsen, P., & Funch, M. (2013). Does Board Diversity Really Matter?: Gender Does Not, but Citizenship Does. *International Journal of Business Science and Applied Management*, *8*(1), 16–27.

Saeed, A., Mukarram, S. S., & Belghitar, Y. (2021). Read between the lines: Board gender diversity, family ownership, and risk-taking in indian high-tech firms. *International Journal of Finance & Economics*, *26*(1), 185–207.

Sanders, W. M. G., & Carpenter, M. A. (1998). INTERNATIONALIZATION AND FIRM GOVERNANCE: THE ROLES OF CEO COMPENSATION, TOP TEAM COMPOSITION, AND BOARD STRUCTURE. *Academy of Management Journal*, *41*(2), 158–178.

Sarhan, A. A., Ntim, C. G., & Al-Najjar, B. (2019). Board diversity, corporate governance, corporate performance, and executive pay. *International Journal of Finance & Economics*, *24*(2), 761–786.

Schmidt, R. H., & Tyrell, M. (1997). Financial systems, corporate finance and corporate governance. *European Financial Management*, *3*(3), 333–361.

Shleifer, A., & Vishny, R. W. (1997). A survey of corporate governance. *The Journal of Finance*, *52*(2), 737–783.

Sila, V., Gonzalez, A., & Hagendorff, J. (2016). Women on board: Does boardroom gender diversity affect firm risk? *Journal of Corporate Finance*, *36*, 26–53.

Singh, V., Terjesen, S., & Vinnicombe, S. (2008). Newly appointed directors in the boardroom:. *European Management Journal*, *26*(1), 48–58.

Smith, K. G., Smith, K. A., Olian, J. D., Sims, H. P., O'Bannon, D. P., & Scully, J. A. (1994). Top management team demography and process: The role of social integration and communication. *Administrative Science Quarterly*, *39*(3), 412.

Smith, N., Smith, V., & Verner, M. (2006). Do women in top management affect firm performance?A panel study of 2,500 Danish firms. *International Journal of Productivity and Performance Management*, *55*(7), 569–593.

Stulz, R. (1988). Managerial control of voting rights. *Journal of Financial Economics*, *20*, 25–54.

Tajfel, H. (1974). Social identity and intergroup behaviour. *Social Science Information*, *13*(2), 65–93.

Tarus, D. K., & Aime, F. (2014). Board demographic diversity, firm performance and strategic change. *Management Research Review*, *37*(12), 1110–1136.

Terjesen, S., Sealy, R., & Singh, V. (2009). Women directors on corporate boards: A review and research agenda. *Corporate Governance: An International Review*, *17*(3), 320–337.

Terjesen, S., & Singh, V. (2008). Female presence on corporate boards: A multi-country study of environmental context. *Journal of Business Ethics*, *83*(1), 55–63.

Tushman, M. L., & Romanelli, E. (1985). Organizational evolution: A metamorphosis model of convergence and reorientation. *Research in Organizational Behavior*, *7*, 171–222.

Useem, M. (1998). Corporate Leadership in a Globalizing Equity Market. *The Academy of Management Executive (1993-2005)*, *12*(4), 43–59.

Van der Walt, N., & Ingley, C. (2003). Board dynamics and the influence of professional background, gender and ethnic diversity of directors. *Corporate Governance*, *11*(3), 218–234.

van Veen, K., & Elbertsen, J. (2008). Governance regimes and nationality diversity in corporate boards: A comparative study of germany, the netherlands and the united kingdom. *Corporate Governance: An International Review*, *16*(5), 386–399.

Webber, S. S., & Donahue, L. M. (2001). Impact of highly and less job-related diversity on work group cohesion and performance: a meta-analysis. *Journal of Management*, *27*(2), 141–162.

Westphal, J. D., & Bednar, M. K. (2005). Pluralistic ignorance in corporate boards and firms' strategic persistence in response to low firm performance. *Administrative Science Quarterly*, *50*, 262–298.

Williams, K. Y., & O'Reilly III, C. A. (1998). Demography and Diversity in Organizations: A Review of 40 Years of Research. *Research in Organizational Behavior*, *20*, 77–140.

Wright, P. M., & Snell, S. A. (1999). Understanding executive diversity: more than meets the eye. *Human Resource Planning*, *22*(2), 49–52.

Yousaf, U. B., Jebran, K., & Wang, M. (2021). *Can board diversity predict the risk of financial distress?* (Corporate Governance: The International Journal of Business in Society, ahead-of-print(ahead-of-print), ahead-of-print.)

Zald, N. M. (1969). The Power and Functions of Boards of Directors: A Theoretical Synthesis. *American journal of Sociology*, *75*(1), 97–111.

Junior Management Science 8(3) (2023) 591-616

Junior Management Science

journal homepage: www.jums.academy

Managing Customer Success: An Evolutionary Process Model for Role Development in SaaS Entrepreneurial Ventures

Lennart Jayasuriya

Technische Universität München

Abstract

In an increasingly SaaS-driven, competitive entrepreneurial ecosystem, retaining customers has become a key challenge to solve for entrepreneurial ventures. Customer success management provides a possible response to this challenge, looking to build a close relationship with customers to ensure a maximum value-added through the sold software solution. This thesis conducts inductive qualitative research based on eight German SaaS entrepreneurial ventures. It showcases the evolution of the customer success management role in three phases from a 360° support towards a trusted advisor. Over three phases, task change from an operational to a more strategic focus, which is connected to changes in the internal collaboration. The results suggest a strong individual impulse to be a necessary condition for customer success management to emerge and evolve. Furthermore, the role development is accompanied by a perspective shift of the own entrepreneurial venture and a continuously iterating definition of customer success. The findings of this thesis highlight important challenges over the course of establishing a customer success management department in an entrepreneurial venture looking to provide theoretical groundwork for future research as well as start-ups investigating the topic.

Keywords: Customer success management; Entrepreneurial ventures; Process model; SaaS; Grounded theory.

1. Introduction

The German entrepreneurial ecosystem is currently propelled by newly founded software start-ups. In 2021 alone more than 3000 start-ups were founded in Germany, of which more than 50% operated in a Software-as-a-Service (SaaS) subscription-based business model (startupdetector, 2022). At the same time venture capital investments in Germany have reached a new peak in 2022 following an 229% increase to 17,3 billion euros in 2021, of which more than 10 billion euros were invested into software, e-commerce or fintech start-ups (Ernst & Young, 2022). The entrepreneurial ecosystem in Germany seems to have reached the "subscription business model era" (Vaidyanathan & Rabago, 2020). Well-known characteristics present in the B2C sector have now transitioned into the B2B sector as easy-implementation and low switching costs allow companies to buy software on a trial basis to evaluate the potential value before committing to a long-term contract, if at all. With new software start-ups emerging, multinational enterprises (MNEs) as well as entrepreneurial ventures find themselves forced to respond to the challenge of increasing competition resulting in higher customer churn risk. With the entrepreneurial ventures being backed up by significant venture capital, a lot of companies may trust the software solution early on by trying out their product and agreeing to pilot projects or time-limited license agreements.

> *"Today we are in the age of the customer, where the customer is king, and the role around which everything else must revolve is the customer success manager." Geoffrey A. Moore (Vaidyanathan & Rabago, 2020, p. 1)*

But how does an entrepreneurial venture convince a customer of the value of the own solution and secure a long-term commitment resulting in recurring revenue? A potential answer to this challenge could lie in the customer success manager, a role that has gained attention over the course of the last years in public press and academic research (Hilton, Hajihashemi, Henderson, & Palmatier, 2020; Hochstein, Rangarajan, Mehta, & Kocher, 2020; Porter &

DOI: https://doi.org/10.5282/jums/v8i3pp591-616

Heppelmann, 2015; Vaidyanathan & Rabago, 2020). Customer success management (CSM) looks to extend existing customer management practices by not only managing the customer, but actively guiding him towards measurable value created through the sold software solution ultimately creating long-term retention and customer satisfaction.

Existing academic research has shown that roles in entrepreneurial venture early on tend to be loosely defined, amorphous and regularly changing (Gaibraith, 1982; Tsouderos, 1955). Over time and growth of a company, roles become more specialized (Tsouderos, 1955), which can be triggered by developing around an employee carrying out the role (Mintzberg, 1997) or as a response to crises (Adizes, 1979; Gaibraith, 1982; Greiner, 1972). The timing when to change a functional role appears to be a highly complex challenge to respond to, as young companies often tend to update their organizational structure too early or too late, resulting in several disadvantages (Chandler, 1962; DeSantola & Gulati, 2017; Greiner, 1972). Nevertheless, "much remains to be learned about how [...] change in organization design comes about and when and why change does (not) arise." (Alexy, Poetz, Puranam, & Reitzig, 2021, p. 2)

By comparing start-ups and scale-ups with customer success departments of different maturity stages through inductive qualitative research this thesis looks to ground new theory how a customer success management department emerges and evolves in an entrepreneurial venture and how the corresponding role of the customer success manager develops over time. The main outcome of this master thesis is an evolutionary process model that finds evidence that a customer success management department in an entrepreneurial venture develops over the course of three phases with fluent transitions. The maturity of the CSM department did not seem to automatically be included in the scaling process of a venture with some companies beginning to investigate the topic substantially later than others. Triggered initially by a strong individual engagement of an existing, ideally customer-facing employee, in the first phase the role can be described as being a 360° support focusing on multiple, often operational tasks. Over time the demand for customer success managers to be staffed on customers increases in the start-up as the first value of the CSM work is recognized in the company. This triggers the second phase, which is focused on scaling the role by standardizing and outsourcing operational tasks to be able to sustainably grow into a small team. In the third stage, the CSM role can be described as a trusted advisor, which serves the most valuable customers and is billed as an extra service to these customers as well. Operational tasks in the last stage are almost completely outsourced from the CSM department as the role focusses and consulting in software use cases and connecting to the strategic goals of the customer.

Furthermore, this thesis suggests relevant metrics that are relevant for a customer success management department to measure the value of its work over time. It highlights a shift in company perception from being a new player on the market towards a category leader that pioneers a whole new software. This shift in perception shows to be an important accompanying development, as it influences the way customer success management presents itself internally and externally.

The evolutionary process model contributes to academic research by showing *how* a specific role develops in an entrepreneurial venture, thus extending and specifying the existing knowledge on role development. Furthermore, the thesis provides novel academic insights by studying customer success management based on the data of multiple cases for the first time. The results look to provide a first qualitative theory to be used for future research that looks to further investigate the role development in entrepreneurial ventures in general or study customer success management as the newest practice in customer management.

Additionally, the theoretical results look to practically contribute to the entrepreneurial ecosystem in Germany by providing a detailed overview on how tasks and organizational collaboration of the CSM department develop over time complementing the process model. Ideally, the outcome of the thesis can serve as a blueprint inspiration for future start-ups looking to sell their product in a SaaS business model approach with a focus on customer success.

2. Theoretical Background

This thesis looks to ground new theory on how the role of the customer success manager develops in SaaS entrepreneurial ventures. This section will provide an overview on existing academic work in the fields of organizational design evolution and growth as well as the way roles develop in entrepreneurial ventures. Furthermore, initial articulations and definitions of customer success and customer success management are summarized to serve as a theoretical ground for the reader to gain an overview on the topic within the context of customer management research.

2.1. Organizational Design and Growth of Entrepreneurial Ventures

Existing studies in organizational design evolution mostly sample large established companies rather than entrepreneurial ventures (Colombo, Rossi-Lamastra, & Matassini, 2016). A research string of classical studies looking to find universal structures and processes applicable to all organizations was ultimately abandoned (DeSantola & Gulati, 2017). Nevertheless, these studies provided first evidence for the complexity of studying the evolution of organizational designs for future research. Another research string focused on observing a life-cycle perspective in organizational design evolution and shifted the focus towards finding certain developmental stages a venture goes through with its structure, which are correlating with the ventures organizational growth (Kazanjian, 1988). However, these studies were constrained by relying on rather small-sized qualitative case studies and similar to the classical studies looked for ubiquitous structures (DeSantola & Gulati, 2017). This research eventually had the impact of proving that entrepreneurial ventures do not behave

in the same way as larger organizations while at the same time emphasizing the relevance of unique company-specific factors during growth (DeSantola & Gulati, 2017).

Start-ups, scale-ups and unicorns studied in this thesis can be grouped using the terminology entrepreneurial ventures. Start-ups have been characterized as young companies with a strong focus on growth and innovation, which significantly differentiates these ventures from small businesses (Carland, Hoy, Boulton, & Carland, 1984) or small and medium-sized enterprises (SMEs). Moreover, definitions often consider them to be younger than 10 years (Bundesverband Deutsche Startups, 2021). A scale-up can best be described as a start-up showing first success in generating revenue, raising venture capital (VC) and growing their team. Unicorns are scale-ups that have a very high company valuation, with the threshold usually being one billion euros. Research indicates that entrepreneurial ventures often show growth rates higher than those of more mature companies (Kirchhof, 1994), while the challenges that they face during their growth phase are substantially different and unique (Bhidé, 2000). It is important to notice that growth in this context can occur among many different categories, while not necessarily in all of them at the same time (DeSantola & Gulati, 2017). Accordingly, increase in revenue or customers served does not always accompany a simultaneous increase in the organizational headcount (Josefy, Kuban, Ireland, & Hitt, 2015), which is reflected in the entrepreneurial ventures studied in this thesis as well.

A substantial number of entrepreneurial ventures fail within the first 5 years (Åstebro, Herz, Nanda, & Weber, 2014; Dahl & Sorenson, 2012). Additionally, numerous challenges arise from the problem of scaling (Eisenmann & Wagonfeld, 2012), especially when responding not only to an increasing amount of activities, but also to a shift in scope of tasks (Chandler & Hikino, 2004). This indicates that entrepreneurial ventures operate in an environment of uncertainty where they are constantly being exposed to new roadblocks to deal with. Moreover, start-ups and scale-ups are perpetually challenged to establish a compromise between being flexible in their organizational design and being efficient in the way they carry out their business since resources are scarce (Chandler & Hikino, 2004). The continuing need for compromises during scaling suggests implications for role developments in new ventures as well (see section 2.2).

Current academic research agrees that an entrepreneurial venture is constantly facing the ambiguous endeavor of deciding for the right level of structure in organizational design. Too much structure can result in stagnating organizational growth, while too little structure is likely to result in chaos and limitless, inefficient improvisation (Davis, Eisenhardt, & Bingham, 2009). Nevertheless, Davis et al. (2009) criticize that essential potentially relevant factors such as time delays or limited attention span are not considered as well as the failure to describe how structure influences efficiency.

DeSantola and Gulati (2017) identify two prevailing narratives in academic research that interlink the organizational design of entrepreneurial ventures with the aspect of organizational growth, which they label the endurance and the change narrative. The first narrative puts forward that the substantial organizational elements of an organization stay relatively stable and are hard to change, even when undergoing pressure to grow (Beckman & Burton, 2008). This suggests that the initial decisions of founders create a lasting legacy on the organizational development of an entrepreneurial venture (Burton & Beckman, 2007). In contrast to the endurance narrative, the change narrative focuses on how the organizational design changes directly relate to growth. Research in the change narrative indicates that the growth-related increase in complexity is showcased in the organizational design of an entrepreneurial venture (Greiner, 1972; Kazanjian, 1988). In comparison to the frequently addressed endurance narrative, research on entrepreneurial ventures has been rather scarce for the change narrative (DeSantola & Gulati, 2017). DeSantola and Gulati (2017) point out that change narrative assumes that the initial organizational design in entrepreneurial ventures is highly informal and variable especially with regards to task organization. This means that the founders and early employees possess a very central role in guiding and impacting roles (Mintzberg, 1997). Furthermore, companies respond to crises such as the sudden leave of an employee or conflict with evolution in organizational design, requiring the entrepreneurial ventures to respond with adjustments in their structures (Adizes, 1979; Gaibraith, 1982; Greiner, 1972).

This thesis looks to follow the suggestion of DeSantola and Gulati (2017) that future research in the field should try to incorporate aspects of both narratives of organizational design evolution and only "bringing the two narratives together will allow researchers to arrive at a more complete understanding of organizational dynamics during growth" (DeSantola & Gulati, 2017, p. 655).

2.2. Role Development in New Ventures

The traditional academic view on role development states that in early organizational stages functional roles are still relatively amorphous and loosely defined (Gaibraith, 1982). As ventures grow and scale, their functional roles gravitate towards becoming more specialized (Tsouderos, 1955). Although experience-based iterations on a venture's business model have been studied, the question when and how roles and the organizational design of an organization change still remains unclear from the current state of academic research (Alexy et al., 2021; DeSantola & Gulati, 2017).

When dealing with iterating functional roles and decision-making structures during scaling, entrepreneurial ventures are likely to struggle in one of two ways (DeSantola & Gulati, 2017). Studying General Motors, DuPont, Standard Oil, and Sears Roebuck at a point where they experienced high growth Chandler (1962) found that these businesses responded to the ambiguous challenge to be both flexible and efficient by not updating the early role structures for too long. This ultimately led to inefficiencies and seems

to be the prevailing scenario for most entrepreneurial ventures as well (DeSantola & Gulati, 2017). In contrast to this observation, updating roles too quickly can also exhaust sometimes resources due to the overall organizational design not fitting the new position (Greiner, 1972). Current academic literature does not answer the question arising from this dilemma: if there is a correct timing of functional role changes and which actions can be taken to mitigate the observed struggles.

The roles in young ventures, for instance start-ups, are characterized by only little role specialization and simplistic organizational decision-making structures (Mintzberg, 1997). In this initial stage Mintzberg (1997) finds that clear division of labor is still absent, and roles remain loosely defined open to change for future venture needs at a given point in time. Regular communication connecting almost all roles with each other is a key factor to keep the loose organizational structure functional (Gaibraith, 1982; Leavitt, 2005).

Existing academic research has focused investigating the need for efficiency in the change of an organizational design. Over the process of growing as an organization tasks become increasingly complex. At a certain point the entrepreneurial venture formalizes these tasks into a specific functional role with a major challenge of this process being the coordination of efforts across several recently established roles (Aldrich & Ruef, 2006; Blau & Schoenherr, 1971). Furthermore, changes of the organizational design in an entrepreneurial venture that affect the scope of a certain role increase the specialization of these roles and foster team changes (Aldrich & Ruef, 2006). According to Aldrich and Ruef (2006), the establishment of an increasing number of functional roles might even result in reduced direct administrational overhead due to economies of scale in supervision functions. At the same time it creates new challenges through the complexity of role differentiation (Blau, 1970). Potential challenges may even be amplified through conflicts between early team members or founders that experience the phase of scaling and its growing pains (Flamholtz, 2016). An example of such a conflict would be realizing that own expectations towards a career in the entrepreneurial venture are in danger of being jeopardized by new employees hired in senior roles over them (Strauss, 1974).

For likely any organization experiencing growth one of the biggest organizational challenges seems to be facing the ambiguity between the need for efficiency through functional roles. Additionally, the problem of increasing complexity through the establishment of differentiated functions as well as integrating new hires into the existing organizational design has to be dealt with. Consistent with this assumption, Ambos and Birkinshaw (2010) state that entrepreneurial ventures' organizational structures are not linear, but show regular iteration, stagnation and relapse over time related to functional role changes.

For start-ups, the initial structure at the time of founding might differ significantly across companies with implications for more mature stages of the venture. As one of the few longitudinal studies in the field Baron, Burton, and Hannan (1996) find that in emerging start-ups founders and early employees shaping the organization already had blueprints for setting up their human resources department. With one of three dimensions being the way to control and coordinate work, the authors find strong statistical evidence for internal consistency in the role development of the human resource department over time. Especially for high-tech start-ups evidence for having a blueprint to set up a successful human resource department and therefore successful employment situation was observed in a study of 200 tech companies (Baron & Hannan, 2002). Although the researchers focused primarily on the human resources department in their study, they provide evidence of distinct logics behind organizing a department implemented early on.

The organizational design evolution of an entrepreneurial venture is not only based on early plans. For instance, individual employees working in an organization at an early stage can also significantly influence the evolution of an organizational design and functional roles. Miner (1990) finds that a venture having idiosyncratic jobs, which developed around the skillset of certain employees rather than a specified job description, can significantly influence the evolution of the venture's organizational design. The author finds that the emergence or ending of an idiosyncratic job often results in shift in functional role development. This impact resembles another unplannable factor in role development an entrepreneurial venture must account for as some central employees might suddenly leave the company and formerly idiosyncratic jobs might transition into more functional roles. In addition to that, stage models of organizational development have found that responding to other crises at an early stage due to coordination struggles result in the specialization of roles and restructuring of functions (Adizes, 1979; Gaibraith, 1982; Greiner, 1972).

2.3. Customer Success Management

As opposed to the terminologies customer relationship management (CRM), customer experience and customer engagement, only recently has the terminology customer success been the focus in popular press. Although academic literature has shown an increasing focus on the first three terminologies, few academic publications explicitly mention customer success so far (Figure 1). Considering the lack of academic research, a "skeptical researcher is left to wonder whether CSM is just the latest management fad, or a valuable innovation in customer management practice" (Hilton et al., 2020, p. 360).

A prevailing question in academic literature is where to situate the terminology customer success within existing customer management practices. The extensively studied practice of customer relationship management mainly focuses on establishing structures to manage the operational efforts arising with each customer, such as software implementation or billing, and store transactional data (Reinartz, Krafft, & Hoyer, 2004). Customer experience extends CRM by modeling and evaluating the customer's transactions in a customer

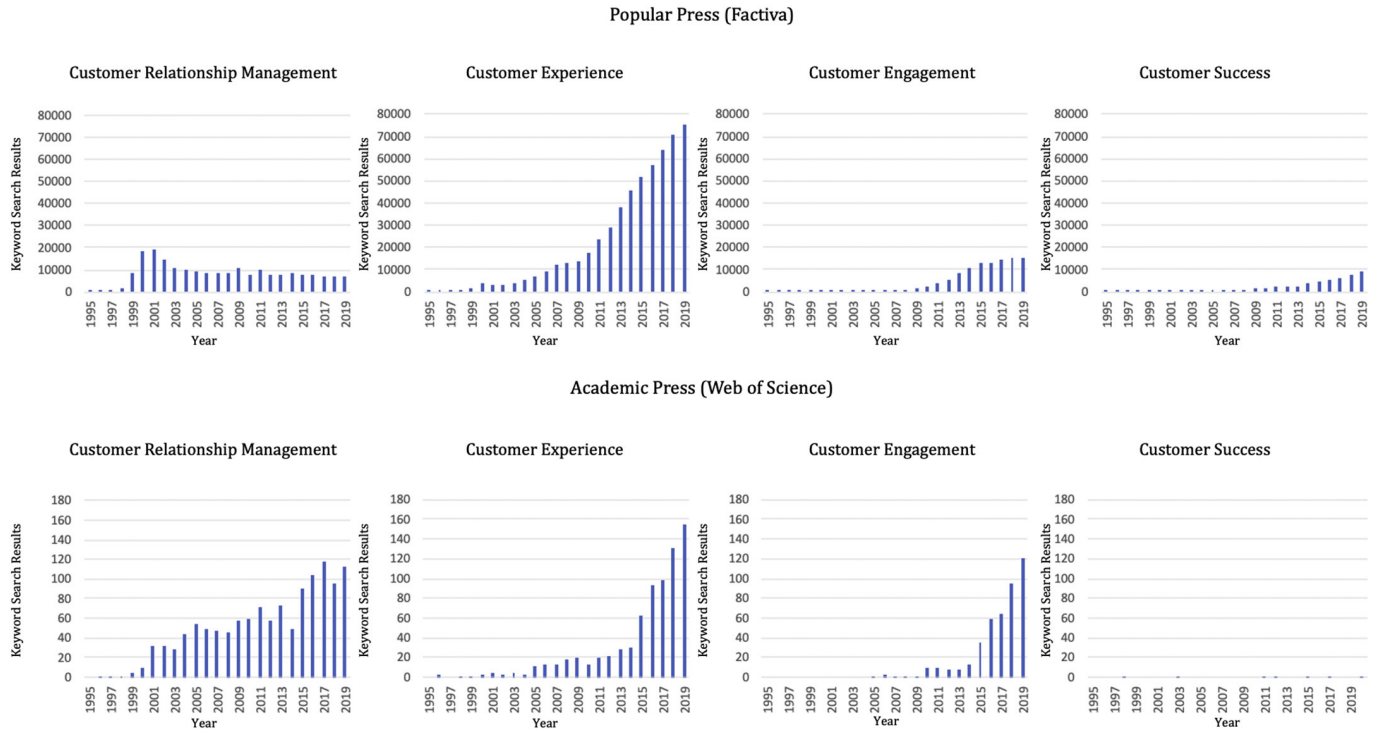


Figure 1: Customer management practice keyword search results by year in popular press and academic press.

Source: Hilton et al. (2020, p. 361).

journey to improve the customer's product experience (Fornell, Rust, & Dekimpe, 2010; Harmeling, Moffett, Arnold, & Carlson, 2017). Customer engagement looks to assess how the customer contributes to marketing functions of the own company, for instance through social media activities (Grönroos, 2011). According to Hilton et al. (2020) customer success management builds on the three practices mentioned. CSM extends CRM by leveraging the transactional data to derive insights on the current health of the customer. CSM extends customer experience by not only considering customer touchpoints with the product, but also assessing strategic or financial goals of the customer. CSM extends customer engagement by not only assessing the customer loyalty, but also considering the customers goals pursued with the own software solution. Therefore, the author label customer success management as "an evolution in customer management practice" (Hilton et al., 2020, p. 368).

A universal definition of the role customer success manager still seems to be absent due to the novelty of the topic customer success itself. Nevertheless, several researchers and business practitioners have voiced their opinion on what customer success is and the objectives a customer success manager should pursue. Vaidyanathan and Rabago (2020, p. 21) define the role as follows: "A Customer Success Manager is the qualified individual that engages with the customer, acutely assesses their needs, strategically aligns the use of your products or services to achieve those needs and ensures that the customer attains their expected outcomes by tactically and proactively taking actions all along the way." (p. 21). The authors propose that the customer success manager is responsible for tackling and overcoming a "consumption gap" (Wood, 2009, p. 1), which results out of the customer's use of the product being lower than the actual capabilities of the product. Closing this gap through the creation of valuable use cases and the incorporation the latest software feature updates is the core task of the customer success manager. Ultimately, a successful execution of customer success should result in churn prevention and help to retain the customer in the long run (Vaidyanathan & Rabago, 2020).

Summarizing the previous research and business articles on the topic customer success, Hilton et al. (2020) situate customer success management as the next evolution of customer management practices and identify leading initial definitions of customer success management across literature. The authors position customer success management between goal management, learning management, and stakeholder management. Accordingly, as a main goal customer success management should improve performance of company and customer.

Zoltners, Sinha, and Lorimer (2019) define CSM as showing the customer the value a solution can provide and how to achieve it. The authors also discuss that companies seem to always go down one of two roads when establishing customer success management: They either rebrand their customer service or account management departments to customer success management departments while keeping former tasks identical or clearly differentiate the role from other customer facing roles, such as sales, through the focus on cus-

tomer success.

Compared to Zoltners et al. (2019), Hochstein et al. (2020) as well as Porter and Heppelmann (2015) provide relationship-focused definitions with an emphasis on managing the customers' experience and the users' engagement with the own solution. While Porter and Heppelmann (2015) see the CSM department in charge of coordinating the efforts of the marketing, sales and service unit of a venture towards benefitting the customer, Hochstein et al. (2020) distinguish the CSM role from marketing, sales and service as an advocate for the customer that makes use of key performance indicators (KPIs), such as for example customer health scores, to retain the customer. At the same time, the authors put forward that all roles in a venture actively contribute to customer success to a certain degree.

Many definitions on customer success and the role of the customer success manager exist and different approaches towards interpreting the role can be identified. However, research on organizational design has identified that roles in entrepreneurial ventures change over time (section 2.2). Accordingly, the research of this thesis looks to tackle the research gap between the initial, rather generalized articulations of customer success management and the fluid development of roles in entrepreneurial ventures. The motivation behind the research gap is that if roles in entrepreneurial ventures change over time, the role of the customer success manager must experience significant changes as well. Moreover, the underlying triggers for role changes and the initial investigation of customer success management are not known leading to the question: When does a customer success management department emerge in an entrepreneurial and how does it evolve?

3. Methodology

The following paragraph will outline the research methodology chosen in this thesis to tackle the research question how and when the customer success management role evolves in SaaS entrepreneurial venture. An inductive qualitative research approach was chosen due to the novelty of the topic customer success management with the goal to create an evolutionary process model for the emergence of customer success management departments in entrepreneurial ventures with a SaaS business model. Data collected from eight German entrepreneurial ventures with a SaaS business model was analyzed in a grounded theory based approach (Eisenhardt, 1989; Gioia, Corley, & Hamilton, 2013; Glaser & Strauss, 1967). After purposefully sampling the cases, semi-structured interviews served as the central input for the inductive data analysis. The gathered data was further enriched by internet research on the start-ups, scale-ups, and unicorns studied. While the initial goal of building grounded theory from cases was pursued, the execution of the research design involved several iterations, which especially focused on the process model development and were the result of regular exchanges between the author and his supervisors.

3.1. Research Design

Customer success management specifically has only been starting to become a focus of academic research in recent years (Hilton et al., 2020) and no former research studying it in entrepreneurial ventures has been conducted before as existing research focuses on customer success management in larger corporates or MNEs. In this thesis, the development of the customer success management role was studied with reference to existing observations of role development in entrepreneurial ventures that have investigated how functional roles develop over time (Aldrich & Ruef, 2006; Baron & Hannan, 2002; Blau & Schoenherr, 1971; Greiner, 1972; Miner, 1990).

As this thesis looks to create a novel insights on how customer success management emerges and evolves in entrepreneurial ventures it grounds new theory based on case studies inspired by Eisenhardt (1989) and Gioia (2014). While initially trying to follow Eisenhardt's (1989) suggestion on how to build theory from cases, it must be clarified that the methodology ultimately incorporated inspirations from different scholars and research practices. The approach combined the grounded theory approach (Corbin & Strauss, 1990; Glaser & Strauss, 1967) with previous work on qualitative research methods (e.g. Miles & Huberman, 1984) as well as case-study based research strategies (Yin, 1981, 2009) and incorporated cross-case analyses (Eisenhardt, 1989). Originating from a social science background (Glaser & Strauss, 1967), grounded theory has become relevant as a research method in business research, specifically in contexts that require new theories (Douglas, 2011). Eisenhardt and Graebner (2007) clarify that the key focus of this research style is to develop a theory rather than testing it, which does not mean that theory building from cases is less objective or precise than large scale hypothesis testing based on random sampling. Building theory from cases focuses on exploring areas of interest that current academic research has not covered yet and connect it to existing academic work (Eisenhardt & Graebner, 2007).

Current research has stated the need for more longitudinal studies in the field of role development in entrepreneurial ventures (DeSantola & Gulati, 2017) with the goal to explain how and when a certain role emerges and develops (Alexy et al., 2021) rather than describing specific stages of an organizational design in a rigid way. The evolutionary process model created in this thesis looks to contribute to this demand by investigating the CSM departments of eight comparable SaaS-based entrepreneurial ventures that differ in maturity with the goal to elaborate how the role develops and evolves over time in several phases. Since the ventures were also asked how processes changed over time, the process model may even be seen as an abstracted version of a longitudinal study placing different comparable companies on a timeline and studying the evolution of their CSM departments in comparison to each other. However, when asking questions that thematize past events, potential retrospective biases of the interviewee have to be accounted for (Cox & Hassard, 2007; Graebner, Martin, & Roundy, 2012). Fur-

thermore, the perspective on changes over time allowed to compare ventures that are already more mature in their CSM department with less mature ventures and find patterns and differences in the development of customer success management across different cases.

3.2. Data Collection and Sampling

For the selection of the entrepreneurial ventures to be studied purposeful sampling was chosen as a technique to explicitly identify and select information rich cases relevant to the research question. The cases had to show homogeneity to be comparable (Patton, 2002), which is why only German SaaS entrepreneurial ventures with an existing customer success department were chosen. Across the cases heterogeneity was observed in the maturity of the customer success management departments of the respective entrepreneurial ventures, which the process model was built on.

Collected data in a grounded theory approach can come from various sources (Corbin & Strauss, 1990). In this thesis the main source of data are semi-structured interviews with employees or founders of start-ups or scale-ups that work in customer facing departments or customer success management. According to Bernard (2006) the willingness to participate and the ability to communicate are important when choosing purposeful sampling, which is why the interviewees got assured anonymity of their name as well as their company to foster information sharing before the interview. The interviews were held in the preferred language of the interviewee, which were English and German for the same reason. Due to the ongoing Covid19 pandemic and geographical distances all interviews were conducted and recorded via video call to be transcribed afterwards as support for the analysis.

The semi-structured interviewing method is a way of data collection well suited for exploring perceptions of different individuals with the opportunity to clarify answers that might not have been fully accurate to the question compared to a standardized interview structure (Barriball & While, 1994). The interview guideline (Appendix A) consists of three main parts. First, descriptive information on the company, its customers and the customer management were gathered. Additionally, a longitudinal view was incorporated by asking how customer management processes in the entrepreneurial venture have changed over time. In a second part, the interviewee was questioned on the own perception of customer success as well as the organizational setup of the customer success department. The objective of this part of the interview was to investigate out if the different cases talk about the same concept when mentioning customer success and a prevailing narrative can be identified. The final part of the interview targeted the interviewee's perception of the market, competition, and the role of the own company within the industry to capture how the interviewee perceives the own market position. After formulating the initial guideline, it was pilot tested at the company of the author to improve the questions and inhibit any potential interviewer bias when conducting the interview as suggested by Chenail (2014). The adjustments are highlighted in the interview guideline (Appendix A).

A total of eight cases (Table 1) was analyzed as data sources in this thesis. Since the six months timeframe of the thesis did not allow for longitudinal study of the cases, but the goal was to create an evolutionary process model, a homogenous sample was created that showed heterogeneity in company maturity. This heterogeneity suggested a potential for the CSM departments of the respective ventures to be also of different maturity. When the samples are relatively homogenous, a sample size between seven to twelve cases is enough to gather sufficient data for inductive analysis (Guest, Bunce, & Johnson, 2006). Following the goal of creating a relatively homogenous sample the entrepreneurial ventures selected had to fulfill the following criteria:

- Founded in Germany,
- founded after 2010,
- offering a software solution,
- SaaS/subscription-based business model,
- having an existing customer success management department or responsible employees for customer success,
- already serving customers with monthly recurring revenue.

In addition to the interview information publicly available information was added to the cases to get information that allows to compare the maturity of the different companies (e.g. number of employees, founding date, VC funding) and enrich the insights given during the interview. Wherever possible the company data was verified with the respective interview partner to ensure validity and accuracy.

3.3. Data Analysis

For the process of data analysis inductive coding was chosen to be fitting with the grounded theory approach of this thesis (Corbin & Strauss, 1990) and carried out as proposed by Gioia et al. (2013). The data analysis process mainly consisted of three steps. As mentioned before, the overall process contained several iteration loops and discussions among the author and his supervisors, where the data was revisited under new considerations.

First, all the interviews were re-read again after transcription performing open coding (Creswell & Plano Clark, 2018; Douglas, 2011) to generate first order concepts that might be relevant for the research question of this thesis. To support the coding progress the software MAXQDA was used. Informational data on the company, which was gathered through the internet and verified during the interview, was added to assess the maturity of the company and its CSM department.

After the open coding was done for all the interviews, axial coding helped to regroup the data to find second order themes, clusters, and relationships between the codes across

Table 1: Case overview.

Company	**Software Solution**	**Interviewee** Role	**Founding** Date	**Venture Capital (in Mio €)**	**Number of Employees**	**Number of Customers**	**CSM Department Size (FTE)**
C1	Hospital Task Management Software	Founder	2018	1,5	15	10-15	3
C2	Gastronomy Optimization Software	Head of CSM	2017	equity	10	30	3
C3	Procurement Software for SMEs	Founder	2020	6,3	20	8	1
C4	Digital Factory Cloud Software	Head of CSM	2013	96	260	100	4
C5	AI chatbot software	VP Customer Strategy	2016	40	120	200	8
C6	Finance operations software for SMEs	Head of CSM	2019	3	45	120	3
C7	HR Management Software for SMEs	Senior CSM Manager	2015	524	1200	6000	23
C8	Process Mining Software	Senior CSM Manager	2011	1400	2000	not disclosed, estimated 4000+ based on internet research	not disclosed, estimated 20+ based on LinkedIn Research

all interviews. This was the first point, where clusters were discussed between the author and his supervisors and iterated after new points came up.

In the third step the clusters were aggregated to higher order themes that could become components of an evolutionary process model. These themes were compared to the academic literature reviewed before to find resemblances with existing academic findings and significant differences, especially in the fields of organizational design evolution and role development in entrepreneurial ventures.

Drawing inspiration from fellow researchers that have designed process models before, the synthesized data was put into a model with three phases.

3.4. Maturity of CSM Departments

As pointed out before, the cases were purposefully selected to try and create a perspective over time to identify different phases an emerging CSM department in an entrepreneurial venture goes through, similar to what a longitudinal study following a single company over the period of creating and developing a customer success department would do. Comparing the different entrepreneurial ventures' organizational designs as well as their CSM departments, the cases were able to be segmented into three different subgroups corresponding to the maturity of their customer success management departments. Based on these subgroups an evolutionary process model was created with a focus on describing the process of how a CSM department emerges and evolves over time in three different phases. Rather than trying to define clearly distinguishable stages every venture goes through, the different phases resemble a fluid concept of a process model, which blend into each other. This means that an entrepreneurial venture can be situated in two phases at the same time while transitioning and several factors have to be fulfilled for the company to completely transition from one phase to another.

The first subgroup consisted of three cases (C1, C2, C3). All entrepreneurial ventures were younger than five years, served less than 50 customers and had between 10 and 20 employees with usually one employee in charge of customer success. For cases C1 and C2 the employee lead two other employees that were mostly focused on operational tasks. The employee or for C1 and C3 a member of the founding team focusses on building up the department, trying to establish first organizational structures and at a later point in time training the first hires in the CSM department. The major focus of the customer success department of this subgroup was "being pushy with customer success" (C3) while "putting out fires" (C2). The customer success managers were heavily engaged in numerous tasks from technical support to training users as well as regular meetings to check in on their customers. At the same time customer success appeared to be a relatively new terminology in the company and while different ideas were present of what the terminology meant, the role and task definition was still very amorphous and reactive to what customers demanded, indicating a rather low maturity of the overall customer success management department in these ventures.

The second subgroup consisted of three cases as well (C4, C5, C6). The ventures existed between five to ten years since their founding, served a three-digit number of customers and had between 50 and 300 employees. The CSM departments headcount was between three and eight employees, with several employees having multiple years of experience in the venture, often in other customer-facing roles such as technical account management or sales. The customer success management department evolved like the first subgroup out of the engagement and ownership of one employee entrusted with setting up organizational structures for customer success. In all cases said employee that had shaped the customer success department obtained a leadership role in the department over time and while still actively managing customers himself, was focused more on internal strategic development of the CSM department. While first structures were already existing, the strategic focus of the ventures in the subgroup was on achieving scalability and standardization of operational processes carried out by the customer success managers. When compared to the first subgroup, the daily operational work from customer success managers focused less on immediate technical support or implementation efforts, but rather on engaging with the customer to find use cases for the software solution, thus acting rather on the advising than the operational facet of the role. Based on these observations a higher degree of CSM maturity was attributed to the cases in this subgroup compared to the first subgroup.

The third subgroup consisted of only two cases (C7, C8), which differed significantly compared to the other two subgroups in size and by being highly valuated unicorns with the most mature CSM department. Despite not being significantly older than the other cases with seven (C7) and eleven (C8) years, the two ventures had already surpassed what could be considered a scale-up phase and turned into a "hyper-growth company" (C8), which is reflected in the company's numbers as well. Both ventures were in the four-digit numbers when it came to headcount (1200 for C7, 2000 for C8). The same statement held true for the numbers of customers served, totaling 6000 for C7 and while C8 did not want to disclose this number, similar numbers can be assumed based on size and internet research on the company. Compared to the other two subgroups the CSM work was strongly focused on connecting to the customer's long-term strategic goals and creating measurable value with the software solution sold. Former obstacles for scalability such as standardization of onboarding workflows, technical support or user education had been outsourced from a customer success manager's perspective. As a consequence, the customer success managers were able to focus on the areas mentioned above with a stronger focus on advising than supporting in operational tasks. Overall, these factors qualified C7 and C8 to have a more mature CSM department compared to the second subgroup. Furthermore, both companies are considered the only unicorns among the cases based on their valuation and VC funding.

The maturity of the customer success departments did not seem to be dependent on company size in terms of total number employees. For instance, C1 and C2 already had three employees dedicated to customer success management despite having 15 or less employees in total. Opposing to that C4 had four employees focusing on customer success with a total headcount of 260 employees, while both companies had focused on establishing customer success management for about one year. This indicates that the growth and the maturity of the CSM department might to a certain degree be detached from the growth of the entrepreneurial venture itself. Just because the venture starts to scale, it will not automatically put a focus on customer success but could for example be strongly guided by the founders in early-stage decision as suggested by (Mintzberg, 1997). C7 and C8 have larger CSM departments than the other cases. However, compared to the total company headcount the CSM departments of the two ventures were the smallest in relative size accounting for only 1,5% of all employees compared to 5% or higher for other entrepreneurial ventures. This could either signal potential obstacles that are yet to be overcome or a point of saturation in terms of internal demand for customer success managers.

4. Results

This section describes the findings of the inductive research of this thesis, which can be summarized in an evolutionary process model that describes how a CSM department in an entrepreneurial venture emerges and develops in three phases (Figure 2). The model is read from left to right and underlined by a timeline of increasing demand for customer success managers to be staffed on customers in the entrepreneurial venture. In the center grey rectangles with arrows resemble the three phases and the describe state of the customer success management department. The white ovals below the phases symbolize challenges during development. Specifically during the transition of phases, they resemble triggers to move into the next phase and thus stretch over a period of time. Above the phases organizational changes accompanying and resulting out of the CSM development are visualized in dotted boxes. Furthermore, the model is traversed by perception changes of role and company represented in dashed lines.

In this section, first the identified initial triggers factoring into the establishment of a CSM department are outlined. Afterwards, the three phases are elaborated in depth, describing the evolution of the role from a 360° support to a trusted advisor and metrics supporting the role throughout the process. For each phase customer success definition, team structure, task focus, internal collaboration, company perception and challenges are analyzed. Over the course of elaborating on the phases, critical success factors for succeeding with customer success management are pointed out within the different categories. Furthermore, examples as well as direct quotes from the interview support the argumentation behind the process model. As a practical implication to contribute to the entrepreneurial ecosystem, Figure 4 provides examples of concrete tasks and collaborations carried out by customer success managers through the different phases. In combination the figures 2 and 4 look to serve as a blueprint for future start-ups wanting to build a customer success management department and connect to the goal of this thesis to create entrepreneurial impact.

Lastly, a prevailing narrative on the terminology "customer success" is defined looking to contribute the existing academic discussion on the meaning of customer success in entrepreneurial ventures.

4.1. Initial Triggers for establishing a CSM Department

A CSM department did not automatically seem to be part of the scaling process of customer management in start-ups at a certain point of time. Consequently, different triggers were conjected to factor into the initial establishment and the development of the role. This means that the customer success management role is and will not be present in every SaaS company, which leads to the question: Under which conditions does an entrepreneurial venture begin to invest resources into customer success management?

The most important prerequisite for setting up a CSM department is having paying customers to serve in the first place, since it "does not make sense to deal with customer success, when you struggle to sell your product to your first customers" (C5) or are still in the process of trying to negotiate with pilot projects for a financial commitment. In addition to that, the software solution of all cases seemed to have product-market fit serving a yet unmet need of the customer or being an industry specialized market leader in a certain category of software.

> *"I kind of jumped in and said, okay, I will take over the customer success management part." (C5)*

Furthermore, the data shows evidence that a strong individual impulse either from the founders themselves (C1,C3) or an individual employee of the sales organization of the start-up (C2, C4, C5, C6) triggered the company investigating customer success. Usually, the individual employee that originated the impulse, had a background in a customer-related team and took over the initial responsibility for investigating customer success. For the cases with more mature CSM departments (C7, C8) the interviewees were not able to recall how CSM was initially established at their company, because they have joined at a stage, where the department was already scaling. In most cases, the individual taking ownership has heard of customer success and had the interest to pursue it in the company. This individual engagement seemed to be important, as for every case in the beginning there was one dedicated employee or founder committing resources into the investigation of customer success under the specific circumstances of the entrepreneurial venture.

The employee impulse for investigating customer success seemed to be the most prevalent trigger for beginning to implement customer success in an entrepreneurial venture.

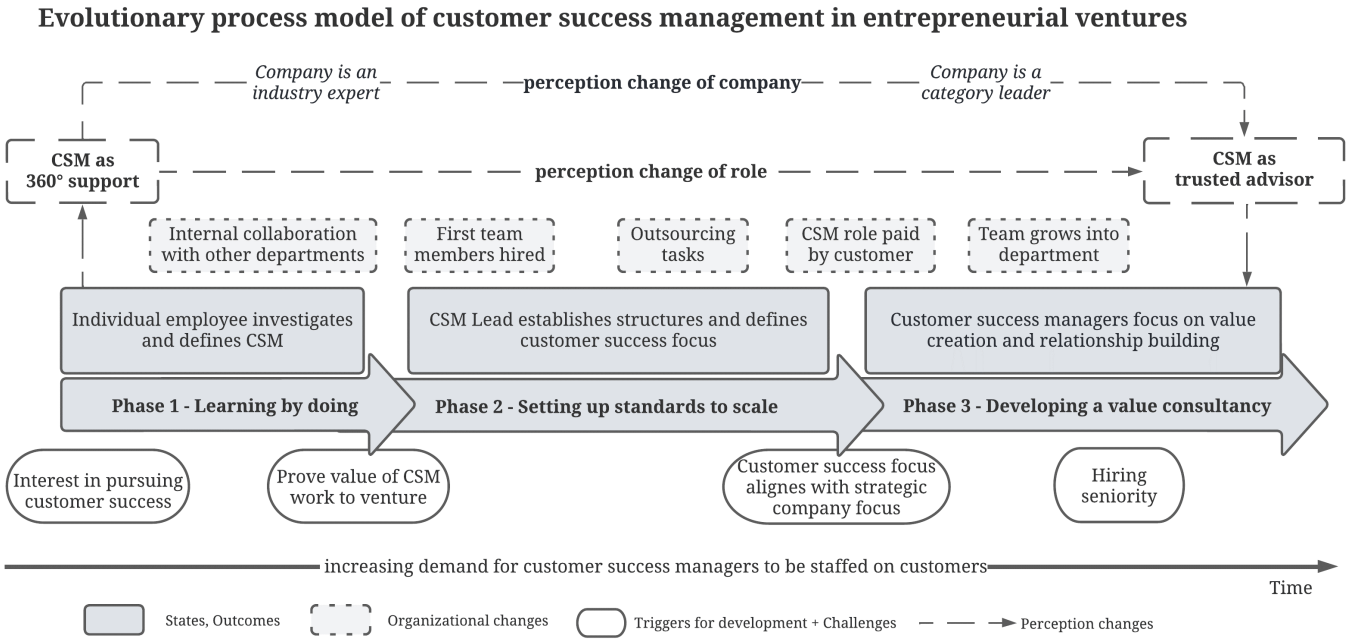


Figure 2: Evolutionary process model of customer success management in SaaS entrepreneurial ventures (own illustration).

Relevant other factors influencing the investigation of customer success mentioned were the own software solution and the customers served. For instance, C2 puts a specific focus on customers in the hospitality industry with a software designed to their customers' needs. The investigation of customer success in this case was triggered by wanting to gain a deeper understanding, if the software solution is providing the value added as intended and be closer to the customer. This company specific focus of customer success indicates a potential reason why there seems to be no generalized one-fits-all approach for setting up a CSM department in a SaaS entrepreneurial venture.

4.2. From 360° Support to Trusted Advisor – Development of a CSM Department

The development of the customer success management department of entrepreneurial ventures can best be portrayed by three phases describing an overarching shift developing the customer success manager from a 360° support to a trusted advisor. The three phases do not have firm boundaries, but rather blend into each other fluently once different triggers or milestones are reached. A supporting overview in which phase the different cases are situated relative to each other can be seen in Figure 3. It is important to notice that the positioning of the entrepreneurial ventures in the model is a subjective snapshot at the time this thesis is written. The CSM departments appear to evolve in a non-linear pace with episodes of rapid growth after overcoming critical challenges followed by a rather levelled growth during the different phases. This makes it difficult to always allocate the entrepreneurial ventures in a concrete phase, with C2 and C5 seeming to be in transition between two phases.

The definition and understanding of what customer success means in the context of the own entrepreneurial venture changed over time. After an initial investigation that leads to a first definition of customer success based on assumptions, the explanation of the terminology grows more sophisticated and distinct over time while being constantly reinterpreted. In the third phase initial statements on the definition may even be revoked as an answer to a change in perspective on customer success and the own company.

The team structure of the department evolved over time starting with a single person with little to no support owning the customer success management role in the first phase. After establishing first structures a small team is built led usually by the employee initially investigating CSM. In this phase all employees are required to strategically contribute to growing the department and must be able to work with little existing structures. In the third phase this team is developed into a department with a double-digit headcount and structures and guidelines to introduce new hires to the customer success manager role.

Over the course of developing the customer success manager role and the department, the tasks in the role description undergo a change as well, which seems to be connected to the change in focus for customer success over time. Overall, the task focus shifts from rather operational tasks that over time will be located in other emerging departments to value-driven, strategic tasks with little to no operational work done by the customer success manager in more mature departments. As certain tasks are discontinued to be a responsibility of the customer success manager it does not necessarily mean that they are no longer necessary or require no employee to deal with due to automation. Accordingly, the

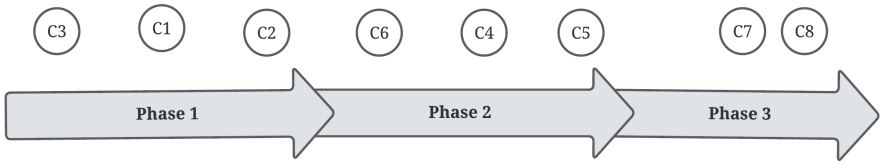


Figure 3: Placement of the sampled CSM departments in different phases of the evolutionary process model (own illustration).

tasks must be relocated to other departments, which might be emerging over the course of the entrepreneurial venture scaling. An overview of the shift in focus of tasks performed by the customer success manager over the course of the department evolving as well as the outsourcing of tasks to other departments can be seen in figure 4.

Additionally, the role of a customer success manager distinguishes itself from other roles in entrepreneurial ventures significantly in the internal collaboration indicated through having task-related connections to almost every department in the entrepreneurial venture. While this might be the nature of things especially in a very early-stage start-up, the ties to other roles seem to persist for the customer success manager even during organizational growth of the venture (Figure 4). With the customer success manager being "one of the biggest knowledge holders in the company" (C4) it is essential to distribute knowledge gained through being close to the customer to relevant departments. Accordingly, this includes the important aspect of product feedback to the product and software development teams, but also feedback loops that might not be as noticeable at first sight. For instance, if the marketing team conducts regular activities like creating press releases or other means of external communications (with the customer), the customer success manager will after at a later stage in time notice how these activities have performed and were perceived at the customer.

Take the first task in figure 4 as an example of the evolution of tasks and collaboration. In the first phase, the customer success manager covers all aspects of technical support, from small bugs to more complex software usage questions. At a certain point, a first-level technical support team is introduced, which is able to respond to bugs and technical maintenance issues across all customers. As a consequence, in the second phase the customer success manager only has to support as a second level support for more complex topics, for example how to concretely use the software given customer-specific circumstances. In the third phase, the technical support task is completely outsourced to a more advanced technical support, which frees capacity for other tasks for the customer success manager. Outsourcing in the context of this thesis is always seen from a CSM point of view. While generally associated with tasks being relocated outside the company, outsourcing in this context means a task leaving the CSM department. In most the cases were relocated to another internal department of the entrepreneurial venture.

Furthermore, the way the own entrepreneurial venture was perceived by the customer success manager changed over time. Initially seen as a new player to disrupt a certain industry, the self-perception develops towards being a category leader of a new software category pursuing goals for the greater good. This development in company perception over time seemed to have a reciprocal effect and interdependencies with the development of the customer success management department, which signals a need to regularly rethink the self-image during the emergence of the CSM role.

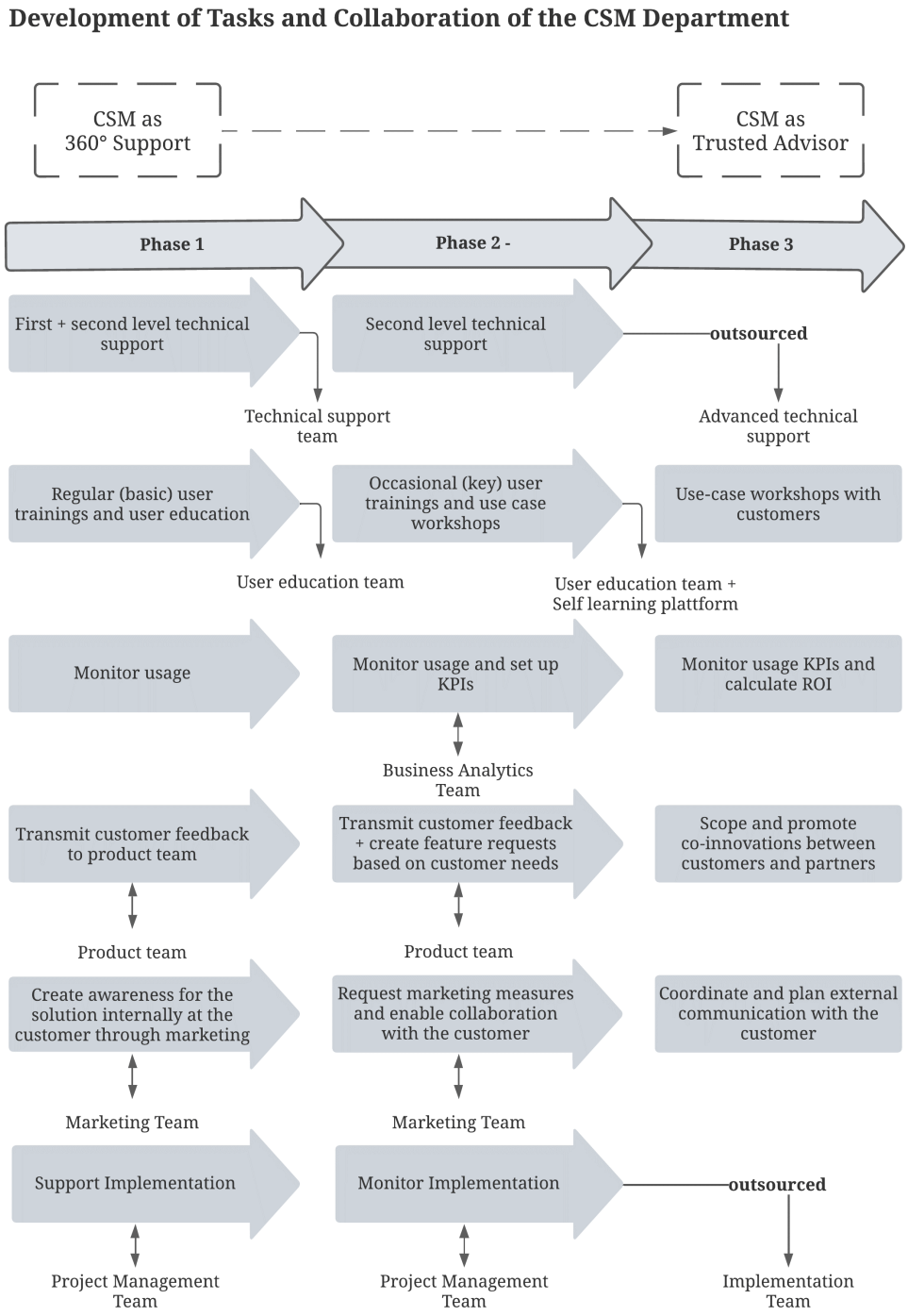
Additionally, over the course of establishing a CSM department challenges were identified, which had to be overcome to transition the next phase. Additionally, these challenges could pose existential blocking points for the role to inhibit growth in customer success tremendously. The time of transition between the phases cannot always be situated to a specific point in time and the importance differs across cases due to a lot of potential influencing variables such as early company structure or available skillsets in the workforce. Hence, the milestones are placed below the phases in the evolutionary process model as triggers to move to the next phase in customer success (Figure 3).

4.2.1. Phase 1 – Learning by doing

After the investigation of customer success is triggered by the initial interest in the topic arising in the entrepreneurial venture, the first phase of developing a customer success department begins. Activities in this phase can be summarized with the label "learning by doing" as the role is characterized by regular changes in tasks and scope and the role description develops quickly with several iterations. The strongest impact in this phase comes from the individual employee owning the customer success topic in the entrepreneurial venture. Often, the responsibility in this phase seems to rest with a single person with only little to no direct support in the form of a subordinate team. As a consequence, a lot of efforts have to be allocated to setting up first structures, responding to various tasks and keeping the workload manageable.

Customer Success Definition

Initially a clear vision and definition of what customer success means in the context is still absent. Consistent with the "learning by doing" label, the employee in charge investigates the topic, most often based on internet research or books. He develops first hypotheses and ideas, how customer success management could look like in the own venture and



Rectangles and grey arrows represent tasks, collaborations with other departments are represented by double-headed arrows and outsourcing of tasks are represented as one-headed arrows

Figure 4: Development of tasks and collaboration of the CSM department over time (own illustration).

what success means with regards to the own software solution.

> *"I think it's supporting and guiding the customer to get the most out of the solution." (C1)*

Accordingly, the initial definition is greatly influenced by the employee in charge of dealing with customer success. At the same time a clear definition of customer success is yet to emerge, which is resembled in rather broad statements, when comparing different answers to the question "How would you define customer success?" across cases. For C1 customer success was about "guiding the customer to get the most out of the solution", signaling a very strong software point of view for customer success. The definition of C3 stands in contrast to that stating that "in the short-term it should be NPS and in the long-term it should be measurable and basically upselling". While this answer underlines the rather vague understanding of customer success, it also shows a more metric and sales focused point of view as opposed to C1. C2 defined customer success very open as "doing everything that makes our customer successful" and "wouldn't like to make this smaller". Apart from attempting a first definition on the terminology, it is crucial that the employee makes efforts to quickly implement interpretations of the definition into the daily operational work. This help to get a first understanding of the customer success managers job description and supports fast iterations for future development.

Team Structure

In the first phase the customer success management department mostly consists of a single person in charge of investigating the topic, hence a team structure with hierarchies and different responsibilities is usually not existing yet. Either the employee in charge is the only person responsible for handling customer success management (C1, C2, C5) or he has one or two employees or working students as support for operational work (C3, C4, C6). In both cases, the initial lead for investigating customer success needs to be highly capable of structuring tasks and working independently to develop a vision for the CSM role. Experience in customer facing roles, such as account management (C2, C4), or sales (C5, C6), is indicated to be helpful when choosing the employee in charge. Furthermore, it seems to be crucial that the employee has a profound understanding of the entrepreneurial venture and the value provided by its software solution, for example by being part of the founding team (C1, C3). Occasionally the impulse to investigate customer success even comes from the employee himself (e.g. C4, C6), which provides another strong indicator for a capable CSM lead that is willing to shape customer success in the entrepreneurial venture in the long term. Quickly it becomes apparent that not all tasks developing in this phase (see 4.2.1.3) can be covered by a single person. This is the reason why the initial lead in this phase already has to plan a team, mostly one to three people, "to get things done correctly" (C3).

Task Focus

The process of developing a customer success department begins as mentioned above with an individual employee owning the at this point often still very abstract topic customer success. Consequently, the employee in charge has a lot of "groundwork to do" (C3) in shaping early organizational structures corresponding to the newly found function. Initially, the customer success manager will cover multiple tasks that are associated with CSM in the company, regardless of having a small team to work with at the beginning (C1, C2, C5) or not (C3, C4, C6). For example, the customer success manager at an early-stage acts as a technical support by supporting the customer with software problems (C1, C4, C6), a trainer by onboarding new users at the customer, a product specialist by transmitting product feedback to the product team and an analyst by monitoring usage of the software (C1). Furthermore, the customer success manager is often asked to support the project management team with the implementation of the software solution at the customer.

Among these tasks, the early focus point of the customer success manager's work seems to differ. While all cases report to cover the tasks mentioned above at some point of their journey, early on most efforts of CSM focused on one specific task, such as "monitoring usage" (C1, C3) or "welcoming and onboarding the customer" (C2). Shortly after, new tasks close to the initial one were added signaling a first uncontrolled development of the role. The reasoning behind this development could be that if you already welcome the customer, you might as well ask him for early product feedback or remind him about the renewal of the license at a later stage. Over time this ultimately leads to a consistent growth in scope of tasks for the customer success manager until a point where changes have to be made to keep the workload on an acceptable level. As the first front line of customer feedback, the position is exceedingly exposed to new demands and might be prone to simply accepting it as part of the own, still very loosely defined job description.

Internal Collaboration

The customer success management role is in close collaboration with several other departments of an entrepreneurial venture. In the initial phase of developing a CSM department, the customer success manager might easily find himself "kind of collaborating with everyone" (C6) as the position is from early on one of the "key knowledge holders" (C4) when it comes to direct customer feedback. This customer centric perspective is for example relevant for the product and R&D team to improve the software solution, for the sales team in negotiations and lead generation or the marketing team to create targeted content for customers. Arguably, the most intense collaboration in this phase appears to be with the sales focused roles of the organization in the form of regular meetings to align on specific customer topics as well as higher-level discussions, which customers to target in the fu-

ture. In some cases, the customer success manager was ideally also already included in presales activities (C5, C7) to foster easier handover. For the other cases there was a handover of the customer from the sales team to the customer success manager after the point of sale.

The nature of the collaboration in the first phase is characterized by task sharing with other departments and assessing customer needs together. As employee hours in the entrepreneurial ventures are scarce resources, the customer success manager often temporarily takes over part of the responsibilities of other departments to enable faster execution. For instance, if an article about the software solution should be published together with a customer in relatively short notice, but the marketing team does not have the personnel capacity to do it within the next two weeks, the CSM can step in and work on the article himself with some support from the marketing department (C4). With the task itself being not even remotely close to the role description, it provides a suitable example for the uncontrolled task development the customer success manager is experiencing.

Company Perception

> *"We are the new kid on the block, in terms of what we do." (C3)*

In the beginning the own company is perceived very focused and specialized. This can be manifested through being an expert in a specific industry (e.g. hospitality for C2), serving a certain type of users within customers (e.g. procurement departments for C3) or a combination of both factors (e.g. everyone connected to production in the automotive industry for C4). Corresponding to this focus, the software solution is rather specialized on the needs and demands of a certain type of customer as opposed to being able to meet demands of several different type of customers. Nevertheless, the own company is classified as a "Software-as-a-service company" (C1) that has a clear vision for how it wants to shape the industry and the market it operates in. This could for instance be "doing everything with performance management and process automatization in the industry" (C2). Despite being confident to create an impact in the market, the cases showed hesitation when calling themselves a market leader due to a yet relatively small customer base.

Furthermore, the ventures did report to have no direct competitors perceived in the market. Although acknowledging that "you never have no competition" (C1), the start-ups attributed themselves a competitive advantage over legacy players in the industry by tackling a digitization gap (C1, C2, C4). For instance, C1 developed the first mobile software in the market as opposed to comparable software solutions that only operate on a desktop. This was perceived as a differentiation and significant advantage over competition increasing the overall confidence in the own product. The confidence in the superiority of the own software seems necessary for the customer success manager to authentically communicate the value added to the customer.

Challenges

With or shortly after the initial decision to investigate customer success, a crucial topic to be dealt with is the differentiation between the customer success manager role and the rest of the existing sales organization. Since a lot of customer success managers tend to already have experience in customer-facing roles, often even in the same company (e.g. C4, C5), they usually already know how to deal with customers and foster retention. It seems in line with the role description of a customer success manager including being close to the customer that the role automatically owns these topics as well. Although this holds true for C5, in most of the cases the customer success manager was intentionally kept out of license negotiations and contract management. These tasks were the responsibilities of account managers or sales representatives while the CSM was still kept up to date through regular exchanges about ongoing customer contract topics. The argumentation was that the customer success manager did already contribute to retaining the customer by carrying out his day-to-day tasks, which work best as a "trusted advisor" (C2) to the customer. Accordingly, the customer success manager add to retention and churn prevention by creating a close relationship with the customer while at the same time pursuing the goal of "becoming indispensable for the customer" (C4). Combining this with the level of trust built up over time, it just becomes "too annoying to leave" (C6) for the customer due to the comfort provided through the work of the customer success manager. This ultimately leads to a greater probability that the customer will sign a license extension when approached by a sales representative without the need of the customer success manager specifically discussing the topic with the customer.

Furthermore, early on a success seemed highly dependent on having "a power user or anyone who will fight for you internally" (C3). Although this is relevant for later stages as well, not having this single point of contact at the customer actively promoting the own software solution might easily turn out to be a blocking point for establishing customer success at the customer. Similar to the position the customer success manager portrays for the customer in the entrepreneurial venture, the point of contact at the customer must advocate internally for the software solution to enable the CSM to support him. This highlights the importance the relationship building has for the work in customer success management. Additionally, when serving larger enterprises as customers, several points of contacts advocating for the solution at the customer can be beneficial.

4.2.2. Phase 2 – Setting up Standards to scale

Having committed resources into investigating customer success at a certain point the value of customer success management has to be proven and recognized in the company itself to prevent potential crises and allow the CSM department to enter the second phase of developing. Customers

are likely to welcome a new role fulfilling their company specific requirements, initially often only at little to no additional costs. However, within the entrepreneurial venture potential conflicts can arise out of the question: Why are we doing this, if it actually costs us money and the customer is already paying for a license? Hence, it is essential the customer success department one the one hand communicates its goals and tasks internally and on the other hand proves the value through tangible results. Potential ways of achieving internal acceptance are communicating improvements company-wide through internal platforms or townhall meetings (C4), supported by tangible evidence such as increased usage numbers, positive customer feedback or improved processes (e.g., easier or faster implementation of the software, faster troubleshooting).

After ensuring the value of customer success management is recognized within the company, similar to the evolutionary process of scaling for an entrepreneurial venture, the CSM department has to prepare for growth and grow out of the 360° support role it has been executing until this point. Consequently, especially the sum of tasks has to become more manageable to be able to easily onboard new employees in the department and furtherly scale up customer success in the entrepreneurial venture.

Customer Success Definition

In the second phase, a first idea of what customer success is has emerged on the one hand through working in the field and on the other hand by being more receptive to input regarding the keywords relevant in the field. Compared to the initial, rather superficial definition, customer success now develops into a more sophisticated concept as interviewees of cases in later phases were able to elaborate more on the topic and differentiate aspects. For instance, C4 reported that for him it is crucial for customer success to build a close relationship with the customer to act as a link between the own venture and the customer. A customer can only be successful if you are "in a regular exchange where the really important topics and developments are discussed" (C4). Hard KPIs, as for example a high user number, are important indicators, but do not necessarily have to indicate success at the customer for him, but always need the personal, close relationship to support it. While C6 extends this relationship-focused definition by focusing on what the customer is trying to achieve with the software solution, C5 creates a contrast to this definition. The interviewee states that through a "lot of handholding" (C5) performed, a successful customer will ultimately try a long-term contract, which signals a rather sales focused definition of CSM. This ambiguity signals once more the impact the CSM lead and the focus of the entrepreneurial venture have on customer success as in this phase the definition becomes more company-specific and might not relate to the literature initially read when investigating customer success anymore. Consequentially, the own definition from the first phase is questioned and iterated based on the goals of the entrepreneurial venture.

What all customer success definitions in this phase had in common, was a focus on enabling the customer to perform best with their software solution and use it to its fullest potential. This corresponded to a shift in task focus and was aligned with the overall idea of scaling the department in this phase.

Team Structure

To be able to scale the CSM department the initial employee investigating the topic needs support and begin to build a leadership role. As the initial value has been proven to the company, a certain degree of financial backup to support the growth of the department exists. If not already done in the first phase, now a first job advertisement for the customer success management position is posted. New hires are handpicked and have to meet high demands, because they have to be able to work with no or partly existent structures in an often fast-growing company. Furthermore, they have to be capable of supporting the CSM lead in scaling the department as a sparring partner for feedback and ideally provide own input on how to grow the department. An ideal candidate to consider would bring experience in customer success management or another customer-facing role into the entrepreneurial venture. Already having industry experience either in the industry of the entrepreneurial venture or the industry of its customers is beneficial too. Additionally, working students or interns can be hired to specifically focus on topics contributing to the growth of the department (C4), such as for example building a KPI tracking for the software or documenting structures. Nevertheless, it can be stated for the German market at the time the data was collected that a scarcity in candidates can be observed with numerous entrepreneurial ventures and established companies looking for customer success managers on platforms such as LinkedIn. The increase in demand for the CSM role Vaidyanathan and Rabago (2020) have pointed out seems to have continued, as every case interviewed was also actively hiring in their customer success department.

As mentioned before, the customer success management lead that usually also set up the CSM department in the first phase should in this phase ideally try to leave operational topics to his newly formed team. This lets him put his focus on strategic topics as well as setting up a plan how to exchange the knowledge on customer success all employees gain within the team. Furthermore, since the CSM lead might suddenly find himself in his first leadership position with staff responsibility, investigating management practices and regularly asking for feedback seem to be important factors as well.

Task Focus

> *"Because we're now into three figures of customers, stuff that might have worked before doesn't really work anymore as it's not scalable." (C6)*

Occasionally unnoticed by himself, the employee in

charge of shaping the CSM department of the entrepreneurial venture might find himself in the process of restructuring the task description of the role rather sooner than later. Having gone through a relatively uncontrolled functional development involving several iterations and negotiations with other departments the CSM role seems to be well equipped with numerous, different tasks. At this stage the employee in charge will likely push for growing the team and hiring first employees as well since a first formal job description should have emerged. Having established first working routines in the department and onboarding concepts for new employees that describe the day-to-day tasks a CSM must fulfill, a first substantial shift in task focus begins to crystallize. Some tasks, such as giving user trainings or technical support are considered to be not part of the customer success management role anymore, but rather other, often newly emerged departments are relied upon to take over this role. For instance, C4 reported that he expects the newly founded user education team to take over trainings in the future that are currently still carried out by customer success managers, which signals a change in perspective on the own role description. Tasks begin to become more abstract, as technical user trainings become use case workshops and the position is described more abstract as "being at the pulse of the customer" (C4) or "deep diving into the customer" (C5). While tasks such as "being responsible for the onboarding cycle" (C6) still persist at more mature stages, other time-consuming operational tasks, such as for example first-level technical support were outsourced as quickly as possible (C4, C5, C7, C8). Contradictory to this observation, the support ticket tasks for C6 persisted despite showing strategic growth of the CSM role over time. Although being one of the earliest tasks to be outsourced for most other cases as it was potentially very time-consuming and easy out outsource, for C6 it resembled part of the central goal for the customer success manager to be the "point of contact for the customer, ideally in a somewhat competent way" (C6). The reasoning shows a rather quantity focused approach to the CSM role with the highest number of customers served per customer success manager at C6 being 67. While the differentiation of C6 easily be seen as a company specific exception, it could also hint at the need for customer success as a general terminology to be loosely defined as it has to suit different company or customer definitions of "success".

Internal Collaboration

Corresponding to shift in task focus, the way the CSM department collaborates with other units of the entrepreneurial venture changes as well. Some tasks are completely handed over to other departments, for others the collaboration changed with the customer success manager leaving the rather operational aspect of the tasks to another role. Furthermore, the customer success manager is latest now established in a dedicated department within the sales organization of the entrepreneurial venture, as it might at an earlier stage still have been part of the product department (e.g. C1). Regular exchanges on a bi-weekly or monthly basis with the product, sales and marketing departments are the customer success manager's opportunity to share information and coordinate efforts across customers together. While sometimes the customer success manager has to still operationally help out, he will in this phase rather request concrete measures from other teams and discuss the execution.

The biggest shift for the customer success manager is that the venture usually has set up some sort of first level technical support in this phase, which the customer or users at the customer can contact for assistance. The support positions will still be in exchange with the CSM department, but customer success managers now focus more on second level technical support for higher level problems. In addition to that, the customer success manager has ideally set up an internal knowledge sharing platform at the customer (e.g., in the intranet), where users can find information such as frequently asked questions or guides how to use the software (C4). This shift in collaboration and carrying out technical support tasks frees up a lot of capacity for the role to strategically work on shaping customer success. In addition to that, software implementation and user trainings are starting to become the responsibility of other departments such as the "implementation team" (C7), "user education" (C4), or an "onboarding team" (C5). Due to the gained experience in these fields, the customer success manager shares his knowledge and supports, but does no longer commit as many resources into these collaborations as opposed to the first phase.

In this phase the collaboration with the product and development team becomes very important. Since the customer success manager is arguably the biggest knowledge holder when it comes to direct, unfiltered customer feedback, it is crucial that he distributes this feedback into the product team as the department will at this stage have an own roadmap how and when to develop the solution. Moreover, out of all customer facing roles the customer success manager is best fit to synergize different customer needs into potential new product feature request and communicate it to the product team. The product team can then factor this information into the product roadmap, ensuring a customer-centric product development in the long run.

Another department emerging in this phase might be a business analytics team, looking to professionalize the tracking and measuring of KPIs. For measuring software usage, the CSM department will closely collaborate with this department and likely also put operational efforts into creating KPIs themselves (C4, C5).

Company Perception

> *"We try to change our perspective now, as we have changed a lot." (C5)*

Self-perception of the entrepreneurial venture undergoes a big change in this phase and should do so in order to pre-

pare the customer success department for a long-term sustainable growth. If not already happened, in this phase the venture lets go of the "start-up" image and defines itself as a "scale-up" (C4). The growth of the own company is recognized through professionalizing and standardizing ways of work, which is also reflected in the CSM department. The entrepreneurial venture itself also begins to see itself pursuing higher goals in addition to increasing revenue and profit, such as for example "trying to become a thought leader in the field" (C5). This perspective change has to be incorporated into the development of the CSM department since it factors into the definition, what makes the company successful. Knowing the answer to this question is the prerequisite for a customer success manager to be able to assess, what makes the customer successful and how the own software solution can contribute.

Regarding market position, the own perception also shifted. The own market experience is well recognized with ventures even attributing themselves "some kind of legacy" (C5) despite being less than ten years old. This self-image shift is accompanied by a high degree of confidence in the own "agile and competent people that want to achieve something" (C4) signaling that the venture is able to set realistic goals due to market experience, but still pursues challenging goals, which should also be the attitude of the CSM department.

Nevertheless, the entrepreneurial ventures at this stage seemed to be very aware of the competition in their field. Despite being confident to be better than competitors, for example by customers switching to the own software solution from a competitor (C5, C6, C7), best practices at other ventures are also recognized. For instance, although considering the customer success work very successful, C6 compared the own tasks to be "not as scale" compared to competitors. Accordingly, being aware of the competition regarding advantages and disadvantages seems to be another critical factor to succeed with customer success in an entrepreneurial venture.

Challenges

Corresponding to the initial challenge in the first phase a major discrepancy among the cases was the organizational boundary between the CSM department and the sales organization of an entrepreneurial venture, which indicates an ongoing challenge in structuring the sales organization. For some cases the customer success manager was (ideally) already included into pre-sales activities while for others there was a strict separation where "sales just focuses on new logos" (C5) and the customer success manager steps in after the point of sales and manages the retention. Another potential way of dealing with the ambiguous responsibility was introducing the role of a "pre-sales consultant" (C4), which was designed to take over the scoping process with the customer and build a bridge between customer success, project management and sales. While the structures might differ according to company specific needs and circumstances, it is important to have a scalable approach to ramp up customer success management for future customers to be targeted and served. A failure to do so could lead to a discrepancy between newly acquired customers and the ability to provide them with customer success management for long-term retention. At the same time, potential crises can arise when negotiating tasks between departments internally, as roles are still rather loosely defined, and no one wants to be overloaded with work.

Lastly, in line with the initial challenge of having a point of contact at the customer that promotes the software solution internally, it has to be defined when to staff a CSM on a customer with which capacity, as likely not all customers can be served, at least not in equal quality. While one might easily opt for staffing customer success managers on the customers with the highest revenue or business potential, disregarding soft factors such as the engagement of the point of contact at the customer can seriously inhibit or even block efforts of the customer success manager (C4). Opposing to that, providing every customer with a customer success manager as C6 does, might result in a less close relationship with the individual customer and the CSM being rather a 360° support role as in the first phase.

4.2.3. Phase 3 – Developing a Value Consultancy

After proving the initial value of the CSM role in the company, to reach the next phase it is crucial that the target and vision of the customer success management department aligns with the long-term strategic goals of the top management of the entrepreneurial venture. For instance, if the goal of the company is to penetrate and grow existing large-scale enterprise customers and capture their business potential, the CSM department should also follow a quality over quantity approach meaning that a single customer success manager likely only serves a handful customers and is thus able to invest more time into the penetration of a single customer (e.g. C4). In contrast to this, a strategy could also be to pursue acquiring as many new customers as possible with every customer having a customer success manager as a single point of contact. Naturally, the quality of the individual work with the customer would have to give way to a more quantitative approach to be able to deal with the number of customers. The actual strategic roadmap of a venture would likely never fully incorporate to be one of the two extremes, but rather a mix or an attenuated version of either one of them. Nevertheless, it is important to spend efforts into defining the CSM strategy, before growing the team from the second phase into the department in this phase, as scalability is a key factor to reach the third phase.

In line with the overall development so far, the third phase of developing customer success management in an entrepreneurial venture is highly focused on generating (measurable) value and takes the longest time to reach. While the terminologies consulting or advising was mentioned by all cases across the three phases, this is the phase where the actual work actually becomes very similar to consulting due to the external perspective and strategic focus the customer success manager has developed.

Customer Success Definition

The definition of what customer success is, and the perception of customer success management undergoes probably the biggest change coming into the third phase. Customer success in this phase redesigns itself with a strong focus on "value". While the terminology itself can have multiple meanings and was mentioned 92 times across all interviews, in this phase the definition of customer success truly reflected the focus on value generation. Not only do customer success managers seem to focus on the most valuable customers in terms of business potential (C7, C8), but all efforts of customer success are streamlined towards creating value at the customer. A big factor influencing this shift is that in this phase customer success management is a role paid by the customer, which is ideally already active during presale activities (C7) or staffed later when the software solution is already implemented. As the customer is now paying not only for the software license but also for the customer success manager, he wants to see a return on his investments, resulting in the focus on creating measurable value. However, a second, more subtle effect seems to accompany this development. As opposed to the former phases, the customer success manager no longer has to convince the customer of the value added through the software solution, but rather switch the focus to implementing said value added with regards to customer-specific needs. This assistance and the expertise of the customer success manager in creating value is what the customer ultimately pays for and expects a return of invest upon. Complementing this focus, customer success is defined as having a "strong partnership" (C7) or "close relationship" (C8) with the customer, shifting the focus even more towards advocating for the customer within the own entrepreneurial venture rather than needing to promote the usage of the own software solution.

For C8 the entrepreneurial venture even chose to rename the position of the customer success manager to "customer value manager" at a certain point in time, showing the focus shift of the customer success definition even in the role description. This signals the importance of constantly iterating customer success terminologies, even if it might seem trivial as in the example of changing the job title by one word.

Team Structure

Starting with an individual growing a small team around a leadership position, the CSM department now really has scaled in terms of headcount with double digit numbers. Although still being relatively small in size when compared to the total number of employees of the entrepreneurial ventures studied (Table 1), there is a need for the department to have team structures as well as onboarding routines established in the second phase. The customer success management lead has now really developed into a managing position and can no longer personally guide and onboard every new hire of the company. Structures have to be set up in the second phase that should incorporate the regular exchanges within the customer success team, with other organizational departments and how to set up a relationship with your customer. As opposed to the second phase, new hires should rather know where to start with their role rather than being asked to support in shaping the role and figure out the daily work of a customer success manager on their own. Nevertheless, the profile the entrepreneurial venture is looking for should still be the profile of a fast-learning generalist, ideally equipped with the ability to quickly understand new, complex technical matters and know how to deal with customers.

Task Focus

As the process of (re)structuring the CSM department continues, the formerly uncontrolled development of tasks becomes mitigated and transitions into a more planned and controlled evolution of the functional role and its responsibilities. Now the focus of customer success management from a task perspective is to try and automate or outsource tasks that limit the scalability of the department due to manual, operational labor of the customer success manager. For example, where customer success managers formerly had to give regular software trainings as live webinars, a user education platform with online self-training can free the role of this recurring operational task (C8). Tasks might also be outsourced, often internally to other departments, such as for example a dedicated first-level technical support department that allows the CSM to only act as a second-level technical support for more complex tasks and rather step into a supervisory, less operational role (e.g., C4). At this point standardized onboarding flows for new customers are well established to further increase scalability over a growing number of accounts. The freed capacity through automation, standardization and outsourcing of tasks is funneled into a strong focus on elaborating concrete software use cases with the customer to create measurable value through the software solution. Furthermore, the monitoring of usage is professionalized and supported through software tools that allow to get measurable insights on usage, for example through cookie tracking (C4).

> *"We act as a trusted advisor for our customer." (C7)*

The process of restructuring the department "ends" with the CSM role becoming increasingly closer to the role of a consultant or advisor for the customer, which is also manifested through the fact that having a customer success manager has now become a service the customer is paying for (C7, C8). This factor has a severe impact on the center of attention for customer success management. The focus of the customer success manager now lies on "creating value at the customer" (C4) and what was earlier classified as operational tasks is now completely absent in the functional scope of the role. Value creation in this more mature stage of a CSM department is for example manifested by thinking the uses cases

mentioned before one step further. While at an earlier stage it might have been sufficient to only establish software use cases for the purpose of achieving a growth in usage, now the use cases must ideally create a measurable return on invest. This demand comes often from the customer side and seems tied to the function of the customer success manager being paid by the customer. It is still essential for customer success to have "these close relationships with their customer" (C8), while it is important to notice that the role now becomes somewhat of a more exclusive service only offered to the customers with a high business potential (C7). This is only possible, because administrative and operational topics such as software implementation, onboarding and training are no longer part of the role description of the customer success manager. Smaller customers are still be served by the software company, but the relationship management is now the responsibility of other departments the CSM department has sometimes little to no strings to.

Internal Collaboration

In the third phase internal outsourcing has happened and the organigram of the entrepreneurial venture now pictures a lot of specialized departments that are focused on different customer management responsibilities, such as for example implementation, onboarding, technical support, or business analytics. The customer success management department is now truly focused on delivering feedbacks to other departments and on having close relationships with their customers. Therefore, collaboration no longer happens on operational topics but rather means exchanging knowledge in regular update meetings. Aligned with the focus on creating value and connecting strategic development of the customer, the customer success manager looks to connect internally with other customer facing roles to find synergies for customers he is responsible for. An example would be partnering up and creating "co-innovations" (C8) between the own software solution implemented at the customer and another software company.

Company Perception

The self-perception of the entrepreneurial venture by the customer success manager in this phase is highly influenced by the growth and reputation of the entrepreneurial venture. The company at this stage is likely to have surpassed the initial scaling phase but is yet still perceived as a "hyper growth company" (C8). At the same time not only the growth dimension is visible in the organizational size, but the solution offered it is now seen as "pioneering a whole new software category" (C8) giving the company the status of a "category leader" (C7). Additionally, companies add the pursuit of higher goals to their strategy such as "giving back to society through a foundation" (C7) or aiming to "build a Silicon Valley company in Munich" (C8). Realizing this shift is relevant for the customer success management department as it comes with opportunities, challenges, and requirements to be considered when serving customers.

As a customer success manager, you can sometimes assume that customers already know about your solution and expectations might be high due to extensive marketing campaigns and well-known, public success stories of implementing the own software solution with measurable value and return on invest. Hence, when working with the customer the focus has to be meeting the expectations of the customer rather than raising them. This emphasizes once more, how important the initial relationship building and maintaining it over time for the customer success manager is and the high requirements demanded from the role. However, in this phase the customer success manager also benefits from the existing use cases by having several blueprints, examples or best practices of established software use cases that are likely to serve as inspiration or even be fully applicable to the new customer, making the job of finding valuable use cases easier.

Challenges

When growing the team from a single digit number of employees into a whole department (C7, C8) a need for seniority in the CSM department among new hires emerges since the initial employee managing the department can no longer provide his experience to every new employee in a sufficient way. This stands in conflict to the fact that the role of the customer success manager is still relatively new and has experienced a severe growth in demand over the last years (Hilton et al., 2020; Vaidyanathan & Rabago, 2020). But how do you hire seniority, where no potential employees with multiple years in the exact position exist? The answer to that question could lie in actively reaching out to generalists. A potential example of a promising profile could be former management consultants that ideally have a background in sales, which, as mentioned above, can be helpful when professionalizing customer success management. Furthermore, hiring from other customer facing roles can provide new input and "a fresh pair of eyes" (C7) for the customer success management department in the entrepreneurial venture. Nevertheless, as MNEs and bigger companies are investigating into customer success management as well and can potentially offer a more attractive compensation with higher job security, the fight for entrepreneurial ventures over candidates for customer success management positions seems unavoidable.

Apart from hiring promising candidates, the biggest challenge is the aforementioned focus on value in this phase. While successful examples of software implementations already exist and might be transferred to other customers, the customer success manager might find himself being suddenly measured by "hard" KPIs as opposed to "softer" measurements, which were able to satisfy customers as well as the own management of the venture before. A CSM might even have put forward statements on feasible use cases at earlier stages that have been slightly exaggerated and could now be called for it. This pressure has to be handled by the individual and amplifies once more the high requirements demanded from the role of a customer success manager in an entrepreneurial venture.

4.2.4. Metrics to measure Customer Success

As mentioned before, across the phases customer success management looks to utilize various metrics to support the own role in day-to-day operations as well as proving the value of the role to the own company and the customer.

Across cases a three-fold approach was able to be observed. Every entrepreneurial venture tracked some form of software usage with industry standard SaaS KPIs, such as DAU (daily active users), MAU (monthly active users) or stickiness (DAU/MAU). Furthermore, well-known customer satisfaction metrics such as CSAT (customer satisfaction score) or NPS (net promoter score) were widely distributed and used. As a third factor, metrics based on software-specific usage extended the general usage metrics, like for example tasks accepted in a clinical task management software (C1) or conversations for an AI chatbot software (C5). This provided insights to the CSM, if the software was used as intended.

More mature departments (e.g. C4, C5, C7, C8) aggregated the metrics named above into a single customer health score, which served as a regular basis for evaluating the situation at the customer and the work of the customer success managers. The most mature departments in the third phase (C7, C8) also had their own CSM work performance measured by monthly recurring revenue (MRR) or return on invest (ROI) calculations, signaling a more revenue-driven than usage-focused approach at later stages.

A prevailing approach when first setting up metrics in a customer success management department with little experience is to orientate seems to be to initially rely on well-known SaaS KPIs for software usage and grow the metrics to be more software-specific and sophisticated with the evolution of the department.

4.3. The Customer Success Narrative

After asking every interviewee the question "How would you define customer success?" one observation became apparent: There seems to be no ubiquitous definition for what customer success management is. A potential reason for that could be that the definition for "success" often seems to be intricately linked to the specifics of the respective entrepreneurial venture and its software solution. While not having an abstract, universal definition for a certain position likely holds true for a lot of other organizational role definitions, there was evidence in the interview for a prevailing narrative that comes close to a fluent definition for customer success.

Most interviewees showed signs that no definition found online or in books about customer success management really fit their own day to day operations, as "it really depends on what you are trying to achieve with your software" (C6). Hence, every interviewee came up with a rather solution-specific, distinct definition of customer success that was based on the own tasks. For instance, the answers included "handholding" (C5), "getting the customer productive on your solution" (C2), "supporting, but especially guiding the customer" (C6), "being an expert" (C4) and "building a close relationship" (C7) or "strategic partnership" (C8). While these quotes resemble only a fraction of the answers, it becomes evident that different associations with the terminology arose for different employees with the same job position. What, however, every answer had in common, was a very customer-centric view that seemed self-evident to the interviewees. The word "customer" was mentioned 30 times across the interviewees' answers to the question and sentences like "of course customer success is about taking a customer perspective and make them happy" (C4) indicate though the casual way of mentioning the customer-centric perspective that this could be a main constituent to the narrative. At first, this observation might seem obvious, since "customer" is part of the terminology "customer success". However, the major difference to other customer related terminologies such as "customer relationship management" seemed to be that achieving customer success is about putting oneself "in the customer's shoes" (C8) and "maximizing the return on invest for the customer" (C5).

> *"The goal of customer success to understand what a customer defines as success." (C8)*

Therefore, the definition of customer success has to include what "success" should mean under the light of the venture's software solution, a task which the customer success lead has to investigate over time and align with the company's goals. For instance, customer success for one venture could mean a profound usage of the software for all customer employees, for another venture you want to enable specific power users to use your solution to the full potential while caring less about the total number of users or penetration of the customer company. In addition to that, the influence of the customer the software was sold to and the user actually using the solution not being identical poses another obstacle when trying to achieve and define customer success. The customer success manager might not have direct access through users or at least has to go through the customer he wants to make successful. While the customer might also likely have the user's success as a priority, it does not necessarily mean he knows what makes the user successful. This opens up the challenge of user success for the customer success manager as well with the goal in mind that if the users of your solution are satisfied you will likely also retain your customer.

5. Discussion

Academic research has found evidence for roles in entrepreneurial ventures to become clearer defined and specialized over time, which is usually reflected through a narrower task focus and more precise functional role descriptions (DeSantola & Gulati, 2017; Miner, 1990; Tsouderos, 1955). As shown by the evolutionary process model after emerging as a variable support role the customer success manager develops into a role that can be compared to an advisor or consultant.

"Customer Success Managers are Generalists." (C4)

The focus and understanding of customer success management becomes increasingly distinct and company-specific over time. However, when observing the development of the tasks performed by customer success managers through the phases (Figure 4), it seems that the role nevertheless requires a broad perspective to be carried out effectively. Moreover, the functional role description grows more precise over time, which complements existing research. However, the customer success management role appears to incorporate a certain ambiguity in its specialization. By becoming more of an advisor than a support role, the specialization gains an abstract dimension compared to other roles as it seems to still require a very generalist employee profile and people "with empathy that are able to understand the customer" (C6). In addition to that, the role must always take a customer perspective and align the expectations with the capabilities of the own software solution. Over time and after changing into a role that is paid for by the customer, the customer success manager seems to become a direct point of contact to the customer's company within the entrepreneurial venture. This signals a perspective shift of the CSM to a much broader level over time and stresses the role shift towards a consultant, which often considered a rather generalist job role.

This thesis supports existing research by finding that the role structure of the CSM role in the beginning is loosely defined and unstructured and grows more distinct and organized over time (Aldrich & Ruef, 2006; Gaibraith, 1982; Leavitt, 2005; Mintzberg, 1997). Key factors are communication and collaboration with other departments and efficiency in task management is important to overcome the challenges of growing the department (Blau & Schoenherr, 1971; Flamholtz, 2016; Gaibraith, 1982). Extending the research of Miner (1990) that roles are highly influenced by individuals in the position, this thesis finds evidence that a high-performing, capable individual could be a prerequisite for the CSM role to emerge. A high number of cases (C1, C2, C3, C4, C5, C6) showed strong individual influences factoring into the definition of customer success and the CSM role with the employee responsible for investigating the role becoming the managing position later on as well.

Referencing the question from Alexy et al. (2021), when and how structural organizational change does occur, the role 'customer success manager' emerges out of a strong intrinsic motivation in the entrepreneurial venture. An additional dedicated individual engagement seems to be the initial trigger for the emergence of a customer success management department. Over time, for a CSM department to successfully evolve from a support-based allrounder to a revenue-generating advising role, further criteria for success appear to be regular questioning and iterating of status quo processes. Stagnation in the development of the CSM department could likely lead to failure, similar to the overall challenges in development of an early-stage entrepreneurial venture (Åstebro et al., 2014; Dahl & Sorenson, 2012).

For the start-ups and scale-ups with more mature CSM departments (C4, C5, C6) the employee that took over the initial ownership role had occupied the team leading management position. This could signal an extension of Miner's (1990) findings that structural change in an organizational design must not necessarily evolve out of predefined goals but can evolve around a certain person fostering the creation of idiosyncratic jobs. In the context of customer success management this dynamic seems to have an additional ambiguous dimension. On the one hand, it seems essential that the individual owning the customer success topic has an employment history in a customer-related position in the company's organization, which held true for all cases. On the other hand, for the second subgroup of cases (C4, C5, C6), where customer success was starting to get more mature, no case existed, where an employee had not been a part of the company for a substantial part of its lifetime since founding. While this could be a characteristic exception due to a small sample size, it could also signal a critical success factor for establishing a CSM department in the first place: an experienced employee that must be retained over the period of growing a CSM department. If the initial employee would leave during the second phase of scaling the CSM department, the growth could potentially be inhibited.

Ultimately, in the third phase the customer success management role focusses a lot on relationship and value. Hilton et al. (2020) ask, if customer success management is the next evolution in customer management practice. In fact, it seems that customer success management is the next evolution of customer relationship management. The definitions of CRM focus largely on getting data based insights into the customer and establishing processes to initiate and maintain customer relationships (Reinartz et al., 2004). CSM takes this definition one step further, by not only maintaining the relationship, but actively engaging with the customer as a trusted advisor. Hence, it is a new role emerging in the field of customer management looking to be defined through future entrepreneurial ventures and academic research. From an internal collaboration point of view, it can be hypothesized that roles will over time rather connect to fewer functions and focus on their specialized functions. For the CSM role the close connections to other departments remained a constant factor over time. However, the tasks accompanying the collaboration shifted from collaborating with the department in operational tasks to rather distributing information and managing stakeholders.

Moreover, Vaidyanathan and Rabago (2020) state that it is the customer success managers task to close the "consumption gap" (Wood, 2009, p. 1) between the customer's knowledge and the potential of a company's product. The observed focus of customer success management on generating measurable value with the software solution at the customer, for instance by calculating ROI or measuring usage, provides support for this statement. Potential ways of closing the consumption gap could be the demonstration and implementation of software use cases at the customer in combination with setting up KPIs that measure value, such as increased

usage or return on invest calculations.

Overall, the results of this thesis provide academic theory on the development of a customer success management department in a SaaS entrepreneurial venture that can also serve as practical inspiration for companies looking to investigate the topic.

6. Limitations and further research

The results of this thesis provide new insights into customer success management in entrepreneurial ventures and showcase a tangible example how a functional role develops over time. Nevertheless, limitations of the research methodology and analysis of the data have to be considered. This chapter will highlight potential limitations and provides suggestions for areas future research could explore building on the results of this thesis.

Although the sampling of the cases was tried to be as homogenous as possible to allow to draw comparison, sampling German SaaS start-ups, scale-ups and unicorns does not represent the whole landscape of entrepreneurial ventures. In addition to that, only entrepreneurial ventures with a customer success management department were studied. While this allowed to ground the evolutionary process model and the findings in this thesis, future research could look into comparing entrepreneurial ventures with and without a CSM department in their development and performance over time. Additionally, customer success management does not necessarily have to be a concept tied exclusively to software products (Vaidyanathan & Rabago, 2020). Despite seeming especially relevant in the SaaS sector, studying CSM exerted in other business model or product contexts could add additional layers to the overall understanding of customer success in general.

Moreover, the need for longitudinal and larger studies DeSantola and Gulati (2017) have pointed out is highlighted once more over the course of this research. The methodology of this thesis tried to incorporate a fabricated longitudinal approach by asking for past developments and choosing heterogeneity in company maturity. While the cases were chosen specifically to fit different maturity levels, studying an entrepreneurial venture that sets up a CSM department over a longer period of time could provide significant further insights into the emergence of the role. As only a snapshot in the company's development was observed, much remains to be learned on how organizational units such as the CSM develop further over time. Although many cases reported the challenges they were or are going through while setting up the department, a retrospective bias of the interviewee must be considered. Furthermore, the path of development was not completely possible to track back for C7 and C8 since the employees interviewed joined at a time when the department already had scaled into the third phase. Therefore, some triggers to reach the last phase and scale the CSM department might not have been captured by the process model proposed by this thesis. Again, a longitudinal study, which ideally accompanies a start-up until it becomes a venture of comparable size to C7 and C8, could provide interesting insights on the success factors of CSM and entrepreneurial venture scaling in general. Furthermore, since the customer success management role is rather young it is impossible to know if potential later phases after the third phase might exist since no several decade old customer success management department exists yet as opposed to other departments of organizations that have been studied longitudinally.

The results of the thesis highlighted the importance of individual employee engagement, when setting up a customer success management in an entrepreneurial venture. Nevertheless, it was only an observation resulting out of cross-case analyses utilizing the interview data. Future research could further investigate this observation and build on research on idiosyncratic jobs (Mintzberg, 1997) emerging in entrepreneurial ventures.

In addition to comparing entrepreneurial ventures with and without a CSM department, another promising research string could be to investigate how and if other customer facing roles, like for example project or account management, may carry out customer success tasks. Companies might actually already perform a variation of customer success management without knowing it and just label it different, as opposed to mislabeling non-CSM roles as customer success manager roles (Zoltners et al., 2019). This would also explain the absence of the customer success terminology in academic work observed by Hilton et al. (2020), as the contents could potentially overlap with existing research under different terminologies.

From the perspective of this thesis future research building on the results could take two main perspectives as starting points – a longitudinal approach studying the impact CSM has in an entrepreneurial venture or the abstract definition of customer success management within the context of customer management practices.

7. Conclusion

As a conclusion, this thesis provides grounded theory on how the customer success management role emerges and evolves in SaaS entrepreneurial ventures based on inductive qualitative research on eight German SaaS start-ups, scale-ups, and unicorns. Existing research has examined the complexity of organizational growth in entrepreneurial ventures, ranging from the pressure of scaling to the challenges of constantly responding to crises by iterating organizational design (Davis et al., 2009; Eisenmann & Wagonfeld, 2012). Additionally, roles in entrepreneurial ventures are characterized by undergoing continuous development and become clearer defined over time (Aldrich & Ruef, 2006; Baron & Hannan, 2002; Tsouderos, 1955). Roles change as a response to crises as well and might be severely influenced by individuals carrying out the functional position (Miner, 1990). Customer success management is an emerging role in customer management, drawing upon existing concepts, such as customer relationship management, customer experience and customer engagement by focusing on enabling the customer to use the

own product to its fullest potential (Hilton et al., 2020; Porter & Heppelmann, 2015; Vaidyanathan & Rabago, 2020).

Combining entrepreneurial role development with the newly emerging role of the customer success manager this thesis finds evidence that the development can be abstracted into an evolutionary process model with three phases (Figure 2). The three phases resemble three subgroups of CSM departments with different maturities identified among the cases and showcase a role development from a 360° support towards a trusted advisor.

Initially, the emergence of a focus on customer success can be attributed to a company interest in the topic alongside a strong engagement by an individual employee, which initiates the first phase. Labeled "learning by doing", this phase is characterized by trying to define a first understanding of customer success while simultaneously managing the emergence of various tasks ranging from technical support to use case consultation. Close collaboration with other departments of the venture characterizes the daily, often operational work of a customer success manager. Although the company is confidently considered to provide a significant value-added with the own software solution, a more concrete long-term estimation of the own market position is yet to emerge in the CSM department. To transition into the next phase, it is crucial to prove the value of the customer success work to the company, for example through increased customer satisfaction rates. The second phase is focused on "setting up standards to scale" (Figure 2). The employee in charge of CSM emerges to a team lead, which hires first employees and aims to set up the department for scalability. Accordingly, he looks to standardize and automate operational tasks or relocate them to other emerging functions. Thus, the internal collaboration is characterized by outsourcing responsibilities internally as a more precise definition of what a customer success manager should develops accompanied by a deeper understanding of the own company. This change in company perception is crucial to transition into the next phase as the CSM strategy has to align with the long-term goals of the entrepreneurial venture. In the third phase, customer success management evolves from a team to a department. Growing in organizational headcount, operational tasks have been outsourced almost completely. This results in the customer success manager's tasks having a more strategic focus, such as initiating co-innovations or measuring return on invest. At this point, a distinct image of customer success has been defined coinciding with a company perception driven by higher goals.

The different phases transition fluently and are triggered by overcoming several challenges for development. Additionally, the evolution of the CSM department is accompanied by a growing internal demand for customer success managers to be staffed on customers. As a consequence, the initially small department must find an approach to not only grow, but scale to meet the newly ascending organizational needs in the long-term.

A narrative around customer success shows that a definition of the terminology cannot be universal but has to consider what success means from the perspective of the own company and the customers. Aligning these perspectives is crucial to achieve a development of customer success management that continuously follows the goal of retaining customers in the long-term.

Discussing the theory with regards to existing academic work, it complements the research finding that roles in entrepreneurial ventures specialize over time. Furthermore, the customer success management role shows an ambiguous variation to this observation. Although the functional role description grows more distinct over time, in the third phase it still requires the employee to have a generalist view due to the necessity to connect with many different internal departments. A deep understanding of the software solution and being able to respond to new contexts are key skills a customer success manager must acquire to fulfill the customer's needs.

Limitations of the methodology have to be considered, as the sample was homogenous and consisted of only eight cases. Although this allowed to inductively ground theory, the findings look to be proved or refuted by future longitudinal studies since such a small sample only allows generalization of results to a certain extent. Furthermore, a comparison with entrepreneurial ventures not focusing on customer success could yield new insights on the factual impact of the role on a venture's performance.

Acknowledging the limitations, I believe that the theory in this thesis provides useful groundwork for future research to build upon and for individuals looking to establish a CSM department in their venture. Formerly, definitions of customer success management mostly relied on philosophical and managerial discussions (Hilton et al., 2020; Hochstein et al., 2020; Porter & Heppelmann, 2015). The evolutionary process model provides a first qualitative theory on the development of customer success management. Furthermore, it signals that the understanding of customer success in entrepreneurial ventures changes over time and differs between companies. This refutes a ubiquitous definition of the terminology due to the fact every venture has to develop a two-folded understanding of success: for itself and for its customers. Combined with figure 4, which shows the development of tasks and the collaboration of the CSM department with other internal functions over time, the process model can hopefully serve as inspiration or a blueprint for future start-ups looking to build or grow their customer success management.

First, I want to thank Christian Dette, Ph.D. for offering me the opportunity to write my master thesis during my time at NavVis, constructively challenging my findings, providing new impulses and generally supporting me throughout the process. In addition, I want to thank my supervisor Prof. Amy Zhao-Ding, PhD from the Entrepreneurship Research Institute at the Technical University of Munich for her fruitful discussions, valuable inputs and the overall excellent academic support throughout my thesis.

References

Adizes, I. (1979). Organizational passages—Diagnosing and treating lifecycle problems of organizations. *Organizational Dynamics*, *8*(1), 3–25.

Aldrich, H., & Ruef, M. (2006). *Organizations evolving* (2nd ed.). Sage.

Alexy, O., Poetz, K., Puranam, P., & Reitzig, M. (2021). Adaptation or Persistence? Emergence and Revision of Organization Designs in New Ventures. *Organization Science*, *32*(6), 1439–1472.

Ambos, T. C., & Birkinshaw, J. (2010). How Do New Ventures Evolve? An Inductive Study of Archetype Changes in Science-Based Ventures. *Organization Science*, *21*(6), 1125–1140.

Åstebro, T., Herz, H., Nanda, R., & Weber, R. A. (2014). Seeking the Roots of Entrepreneurship: Insights from Behavioral Economics. *Journal of Economic Perspectives*, *28*(3), 49–70.

Baron, J. N., Burton, M. D., & Hannan, M. T. (1996). The Road Taken: Origins and Evolution of Employment Systems in Emerging Companies. *Industrial and Corporate Change*, *5*(2), 239–275.

Baron, J. N., & Hannan, M. T. (2002). Organizational Blueprints for Success in High-Tech Start-Ups: Lessons from the Stanford Project on Emerging Companies. *California Management Review*, *44*(3), 8–36.

Barriball, K. L., & While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *Journal of advanced nursing*, *19*(2), 328–335.

Beckman, C. M., & Burton, M. D. (2008). Founding the Future: Path Dependence in the Evolution of Top Management Teams from Founding to IPO.

Bernard, H. R. (2006). *Research methods in anthropology: Qualitative and quantitative approaches* (4th ed.). Lanham, MD, AltaMira Press.

Bhidé, A. (2000). *The origin and evolution of new businesses*. New York, Oxford University Press.

Blau, P. M. (1970). A Formal Theory of Differentiation in Organizations. *American Sociological Review*, *35*(2), 201.

Blau, P. M., & Schoenherr, R. A. (1971). *The structure of organizations*. New York, Basic Books.

Bundesverband Deutsche Startups. (2021). *Deutscher Startup Monitor 2021*. https://startupverband.de/fileadmin/startupverband/mediaarchiv/research/dsm/dsm_2021.pdf. (Accessed: 04.07.2022)

Burton, M. D., & Beckman, C. M. (2007). Leaving a Legacy: Position Imprints and Successor Turnover in Young Firms. *American Sociological Review*, *72*(2), 239–266.

Carland, J. W., Hoy, F., Boulton, W. R., & Carland, J. A. C. (1984). Differentiating Entrepreneurs from Small Business Owners: A Conceptualization. *Academy of Management Review*, *9*(2), 354–359.

Chandler, A. D. (1962). *Strategy and structure: Chapters in the history of the industrial enterprise* (24th ed.). Cambridge, Mass., MIT Press.

Chandler, A. D., & Hikino, T. (2004). *Scale and scope: The dynamics of industrial capitalism* (7th ed.). Cambridge, Mass., The Belknap Press of Harvard University Press.

Chenail, R. (2014). Ten Steps for Conceptualizing and Conducting Qualitative Research Studies in a Pragmatically Curious Manner. *The Qualitative Report*, *16*(6), 1715–1732.

Colombo, M. G., Rossi-Lamastra, C., & Matassini, B. (2016). Entrepreneurship Education: A Selective Examination of The Literature. *Foundations and Trends® in Entrepreneurship*, *11*(6), 427–523.

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

Cox, J. W., & Hassard, J. (2007). Ties to the Past in Organization Research: A Comparative Analysis of Retrospective Methods. *Organization*, *14*(4), 475–497.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Los Angeles and Thousand Oaks, California, Sage.

Dahl, M. S., & Sorenson, O. (2012). Home Sweet Home: Entrepreneurs' Location Choices and the Performance of Their Ventures. *Management Science*, *58*(6), 1059–1071.

Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2009). Optimal Structure, Market Dynamism, and the Strategy of Simple Rules. *Administrative Science Quarterly*, *54*(3), 413–452.

DeSantola, A., & Gulati, R. (2017). Scaling: Organizing and Growth in Entrepreneurial Ventures. *Academy of Management Annals*, *11*(2), 640–668.

Douglas, D. (2011). Inductive Theory Generation: A Grounded Approach to Business Inquiry. *Electronic Journal of Business Research Methods*, *2*.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532.

Eisenhardt, K. M., & Graebner, M. E. (2007). Theory Building From Cases: Opportunities And Challenges. *Academy of Management Journal*, *50*(1), 25–32.

Eisenmann, T. R., & Wagonfeld, A. B. (2012). Scaling a Startup: People and Organizational Issues. *Harvard Business Publishing*.

Ernst & Young. (2022). *Startup-Barometer Deutschland.* https://assets.ey.com/content/dam/ey-sites/ey-com/de_de/news/2022/01/ey-startup-barometer-2022.pdf. (Accessed: 05.07.2022)

Flamholtz, E. (2016). *Growing pains: Building sustainably successful organizations* (5th ed.). Hoboken, New Jersey, John Wiley & Sons Inc.

Fornell, C., Rust, R. T., & Dekimpe, M. G. (2010). The Effect of Customer Satisfaction on Consumer Spending Growth. *Journal of Marketing Research*, *47*(1), 28–35.

Gaibraith, J. (1982). The Stages of Growth. *Journal of Business Strategy*, *3*(1), 70–79.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*(1), 15–31.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. London and New York, Routledge.

Graebner, M. E., Martin, J. A., & Roundy, P. T. (2012). Qualitative data: Cooking without a recipe. *Strategic Organization*, *10*(3), 276–284.

Greiner, L. E. (1972). Evolution and revolution as organizations grow. *Harvard Business Review*, *50*, 37–46.

Grönroos, C. (2011). A service perspective on business relationships: The value creation, interaction and marketing interface. *Industrial Marketing Management*, *40*(2), 240–247.

Guest, G., Bunce, A., & Johnson, L. (2006). How Many Interviews Are Enough? *Field Methods*, *18*(1), 59–82.

Harmeling, C. M., Moffett, J. W., Arnold, M. J., & Carlson, B. D. (2017). Toward a theory of customer engagement marketing. *Journal of the Academy of Marketing Science*, *45*(3), 312–335.

Hilton, B., Hajihashemi, B., Henderson, C. M., & Palmatier, R. W. (2020). Customer Success Management: The next evolution in customer management practice? *Industrial Marketing Management*, *90*, 360–369.

Hochstein, B., Rangarajan, D., Mehta, N., & Kocher, D. (2020). An Industry/Academic Perspective on Customer Success Management. *Journal of Service Research*, *23*(1), 3–7.

Josefy, M., Kuban, S., Ireland, R. D., & Hitt, M. A. (2015). All Things Great and Small: Organizational Size, Boundaries of the Firm, and a Changing Environment. *Academy of Management Annals*, *9*(1), 715–802.

Kazanjian, R. K. (1988). Relation of dominant problems to stages of growth in technology-based new ventures. *Academy of Management Journal*, *31*(2), 257–279.

Kirchhof, B. A. (1994). *Entrepreneurship and capitalism: The economics of business firm formation and growth*. Westport Conn., Praeger. (Praeger studies in American industry)

Leavitt, H. J. (2005). *Top down: Why hierarchies are here to stay and how to manage them more effectively*. Boston, Mass., Harvard Business School Press.

Miles, M. B., & Huberman, A. M. (1984). Drawing Valid Meaning from Qualitative Data: Toward a Shared Craft. *Educational Researcher*, *13*(5), 20–30.

Miner, A. S. (1990). Structural Evolution Through Idiosyncratic Jobs: The Potential for Unplanned Learning. *Organization Science*, *1*(2), 195–210.

Mintzberg, H. (1997). *The structuring of organizations: A synthesis of the research* (21st ed.). Englewood Cliffs, N.J., Prentice-Hall. (The theory of management policy series)

Patton, M. Q. (2002). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). Thousand Oaks, California, SAGE Publications, Inc.

Porter, M. E., & Heppelmann, J. E. (2015). How smart, connected products

are transforming companies. *Harvard Business Review*, *93*(10), 96–114.

Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). The Customer Relationship Management Process: Its Measurement and Impact on Performance. *Journal of Marketing Research*, *41*(3), 293–305.

startupdetector. (2022). *startupdetector Report 2021*. https://startupdetector.de/reports/startupdetector-report-2021.pdf. (Accessed: 05.07.2022)

Strauss, G. (1974). Adolescence in organization growth: Problems, pains, possibilities. *Organizational Dynamics*, *2*(4), 3–17.

Tsouderos, J. E. (1955). Organizational Change in Terms of a Series of Selected Variables. *American Sociological Review*, *20*(2), 206.

Vaidyanathan, A., & Rabago, R. (2020). *The Customer Success Professional's Handbook: How to Thrive in One of the World's Fastest Growing Careers–While Driving Growth For Your Company* (1st ed.). Wiley.

Wood, J. B. (2009). *Complexity avalanche: Overcoming the threat to technology adoption*. Point B.

Yin, R. K. (1981). The Case Study as a Serious Research Strategy. *Knowledge*, *3*(1), 97–114.

Yin, R. K. (2009). *Case study research: Design and methods* (4th ed., Vol. 5). Los Angeles, Calif., Sage. (Applied social research methods series)

Zoltners, A. A., Sinha, P. K., & Lorimer, S. E. (2019). What is a Customer Success Manager? *Harvard Business Review*.

Junior Management Science 8(3) (2023) 617-633

Junior Management Science

journal homepage: www.jums.academy

The hidden potential of boredom – How does the relative perception of boredom influence concentration and task performance?

Joshua W. Coen

Universität zu Köln

Abstract

As boredom in the workplace is an important factor with widespread consequences, research interest in this topic is very high. This paper proposes a way of decreasing workplace boredom and thereby the negative effects associated with it. Based on literature about contrast effects, this paper develops the hypothesis that contrast between boredom levels leads to higher or lower evaluations of task boredom in a subsequent cognitive task. In an online experiment, this theory was tested. The results show that the same task is perceived as less boring when preceded by a monotonous and unchallenging task. An opposite effect for highly interesting tasks could not be determined. In addition, it was found that the boredom induction is linked to significantly lower concentration and that both boredom and inattention decrease cognitive performance. As proposed, the performance influence of boredom transcended tasks, with task C boredom fully mediating the relationship between task 1 boredom and cognitive performance. Further insights and implications are discussed, including a possible bidirectional causal relationship between boredom and inattention. The results implicate that boredom can be deliberately influenced through contextual cues and task order to mitigate its disadvantages.

Keywords: Boredom; Productivity; Work design; Concentration; Contrast effects.

1. Introduction

Boredom is unpleasant, but by no means not a rare state. For western societies, boredom was even claimed "when college students in the United States are polled about their concerns and problems, money is on the top of the list and boredom is number two" (Mael & Jex, 2015). Bertrand Russel even joked that "half the sins of mankind" are owed to a fear of boredom (Russell, 1932). Especially the constant overstimulation through media and technology are assumed to increase boredom both in and outside the workplace, resulting in a "national attention deficit disorder" (Mael & Jex, 2015, p. 144). One study showed that students use their smartphone more when bored at work. However, once they put it down, they were even more bored with their work than before and quickly picked it up again (Dora, Hooff, Geurts, Kompier, & Bijleveld, 2021, pp. 4, 8). Aiming to control boredom, this paper explores whether a task be perceived as more boring if it is preceded by a highly interesting task and vice versa and whether these contrast effects mitigate any disadvantages of boredom.

Especially in the workplace boredom has become prevalent in the last decade. A 2016 study found 43% of workers to be bored and disengaged at work regularly, 52% even for the majority of their work week. Especially younger workers are affected by boredom, and those who experience boredom are twice as likely to leave their company in the short run (Udemy, 2016, pp. 1, 3). Other negative consequences of boredom at work job dissatisfaction (Reijseger et al., 2013, pp. 516–518), depression (van Hooff & van Hooft, 2014, p. 353), health problems (Harju, Hakanen, & Schaufeli, 2014 theft or abuse) (Bruursema, Kessler, & Spector, 2011, pp. 100–102; Metin, Taris, & Peeters, 2016, pp. 260–261; Spector et al., 2006, p. 455), low organizational commitment and turnover intention (Reijseger et al., 2013, pp. 516–518).

A study by Wilson et al. (2014, p. 76) showed that people would rather give themselves electric shocks than be bored with nothing to do – even though earlier in the study, they were willing to pay money in order to stop the shocks. This is a first indication that sensations are viewed differently in different contexts. The concept of psychological contrast is well-documented for other sensations, like intrinsic motiva-

DOI: https://doi.org/10.5282/jums/v8i3pp617-633

tion (Shin & Grant, 2019) or happiness (Brickman, Coates, & Janoff-Bulman, 1978), and in contrast to boredom, creativity (Agnoli, Vanucci, Pelagatti, & Corazza, 2018, p. 46ff; Mann & Cadman, 2014, p. 17ff; Preiss, Cosmelli, Grau, & Ortiz, 2016, p. 6) and convergent thinking (Gasper & Middlewood, 2014, p. 54) increase. Applying these insights on boredom, this paper explores whether the perception and strength of boredom can be influenced through the context it appears in. Based on gained insights, this paper proposes a new approach to prioritize tasks according to their boringness in order to minimize boredom through contrast effects and maximize performance among workflows.

The main goal of this thesis is to explore influence of task order on boredom empirically. Uniquely, this paper attempts to isolate boredom from other influence factors, allowing for specific exploration of boredom effects. In the first part of this paper, the definitions and causes of boredom are addressed, drawing from psychology and organizational research in order to formulate nuanced and informed hypotheses. To explore this, an experiment was conducted, the methodology of which will be derived in Chapter 3. After analyzing and interpreting the results, possible limitations are discussed. To conclude the paper, implications for HR and managerial practice as well as further research approaches are proposed.

2. Background

The following chapter will explore thoroughly the different types and causes of boredom (Chapter 2.1), before turning towards the role of boredom in organizational studies (Chapter 2.2). The background is relevant for hypothesis development, methodology of the experiment and possible implications of the results. Afterwards, concepts of psychological contrast will be discussed and transferred onto the sensation of boredom (Chapter 2.3). Finally, the insights will be utilized in order to deduct multiple hypotheses (Chapter 2.4).

2.1. Definitions and dimensions of boredom

As an emotion, boredom is easy to identify, but it remains "a complex, difficult to define construct" (Goldberg, Eastwood, LaGuardia, & Danckert, 2011, p. 649). One of the first definitions goes back to 1903, when psychologist Theodor Lipps described boredom as "a feeling of unpleasure arising out of a conflict between a need for intense mental activity and lack of incitement to it, or inability to be incited" (Eastwood, Frischen, Fenske, & Smilek, 2012, p. 483f; Lipps, 1903, p. 278). Subsequent scholars (e.g. Greenson, 1953) build on this definition by adding that it is a passive state and people suffering from it are unable to define their desire. The defining feature of most definitions comes down to the suboptimal fit between current activity or cognitive requirements and desired activity or cognitive capacity.

Because the optimal fit could either be over- or underreached, there are two separate directions of boredom (Fisher, 1993, pp. 6–7). Qualitative underload describes the phenomenon when boredom is caused by overly simple, undemanding tasks where people underutilize their mental capacity or skills. An example for this could be a monotonous task like copying numbers. The opposite is boredom through qualitative overload, where people are overwhelmed by an overly hard task. An example could be reading an extraordinarily complicated piece of scholarly literature on a topic that the reader is unfamiliar with. Of course, both of these examples are fully subjective – what one person regards as dull might be relaxing to another. While characteristics like repetitiveness or monotony are related to boredom for most people (Loukidou, Loan-Clarke, & Daniels, 2009, p. 8f; O'Hanlon, 1981, p. 54), boringness is not inherent to any tasks and ultimately remains a subjective criterion decided by person-situation fit, or person-environment fit (Fisher, 1993, pp. 14–15).

Another distinction when it comes to boredom are state boredom and trait boredom (Loukidou et al., 2009, p. 7; Watson, Clark, & Carey, 1988, p. 347). State boredom refers to a transient, temporary experience of boredom during an activity, while trait boredom is an enduring, characteristic shown by individuals over a long period of time. These distinguishable affects are also known under the terms episodic and chronic boredom (Mael & Jex, 2015, p. 136), task-related and personality-related boredom (Haager, Kuhbandner, & Pekrun, 2018, pp. 2, 8) or boredom and boredom proneness (Drory, 1982, p. 144). Whenever the term boredom is used in this paper, it refers to state boredom unless specified otherwise. While different constructs can be classified as the opposite of boredom (e.g. fun, relaxation, enjoyment), this paper focusses on interest as the opposite sensation of boredom and uses the term accordingly (Hamilton, Haier, & Buchsbaum, 1984, pp. 184, 191).

There is no shortage of disagreements within boredom research. There is disagreement on whether boredom should be classified as an emotion, state or trait, and sometimes even whether it is affective or non-affective at all (Westgate & Steidle, 2020, p. 2ff.). However, the vast majority of researchers agree that boredom is negative in affect (Merrifield & Danckert, 2014, p. 481; van Tilburg & Igou, 2017, p. 309; Westgate & Steidle, 2020, p. 2). Similarly, researchers are divided whether boredom is low or high in arousal (Martin, Sadlo, & Stew, 2006, p. 196; Merrifield & Danckert, 2014, p. 481; van Tilburg & Igou, 2017, p. 317). Some researchers, however, argue that low and high arousal are not inconsistent, but the result of self-stimulation, and can thus both appear as the result of boredom (Fahlman, Mercer-Lynn, Flora, & Eastwood, 2013, p. 69). Empirically comparing boredom to other emotions, C. A. Smith and Ellsworth (1985, p. 826) distinguished it as the only emotion that is unpleasant but does not require effort. Another empirical study by van Tilburg and Igou (2017, p. 313) showed that there are virtually no significant correlations to other negative emotions.

Theories on boredom are often based the extensively validated MAC model (Meaning and Attentional Components model) (Westgate & Wilson, 2018, pp. 693–696). According to this model, boredom forms along the two dimensions

meaning and attention. The meaning component refers to whether the current activity fits with a person's goals and values, while the attentional component refers to the congruence of cognitive demands and cognitive resources. If the congruence is not given, attention is either not engaged at all or requires a conscious effort. Westgate and Wilson (2018, p. 693) paraphrase the two dimensions as willingness and ability to engage attention in a given activity. In line with Fisher (1993, pp. 6–7), the attentional component of the theory allows for both qualitative over- and underload (Westgate & Wilson, 2018, p. 695). Notably, the model allows for different profiles of boredom – meaning that either the meaning component, attentional component or any combination of the two can lead to boredom (Westgate & Wilson, 2018, p. 696).

The typology of boredom is important to understand the mechanisms behind boredom and thus to create a valid experiment. Additionally, the assessment can be utilized to define the precise scope of the experiment, i.e. focusing on state boredom. As boredom is a very broad and complex phenomenon, it is important to differentiate between types of boredom, both to manipulate and measure the outcomes related to boredom accurately.

2.2. Boredom in an organizational context

While the effects of boredom might be enhanced through overstimulation and technology usage, workplace boredom is hardly a new problem. Already in a 1978 study, up to 56% of workers expressed that they found their entire job boring, while 79-87% reported occasionally feeling bored at the job. As "boreouts" become more and more common these days (Lufkin, 2021) and workplace boredom is closely related to a number of counterproductive outcomes, research on the topic is highly relevant to practice. And it promises to become even more important: By 2025, it is expected that millennials make up 75% of the global workforce (Key Statistics about Milennials in the Workforce Firstup.io, 2021; Winograd & Hais, 2014, p. 2). This group is twice as susceptible to boredom (Udemy, 2016, p. 3), and 64% of them reported that they "would rather make 40,000 a year at a job they love than 100,000 a year at a job they think is boring" (White, 2014). Especially in organizations, boredom should thus be understood as a self-regulatory state and an "imperative towards meaning" (Barbalet, 1999, p. 633; Johnsen, 2016, p. 1410).

Note that workplace boredom refers to the frequency of state boredom at work. This is different from trait boredom in that the boredom is transient and does not necessarily perpetuate outside of work (Mael & Jex, 2015, p. 139). Precursors of boredom at work include both job and personal characteristics. Looking at job characteristics, repetitiveness and monotony are traditionally seen as major causes of boredom (Fisher, 1993, p. 6). Interestingly, workplace boredom has increased over the last decades, even though monotonous work has widely been automated or replaced (Mael & Jex, 2015, p. 142), which suggests that other causes exist. Next to the tasks themselves, low job resources and demands have been linked to workplace boredom (Metin et al., 2016, pp. 261–262), as have uncommunicative or absent coworkers, as socializing with coworkers is an important source of job satisfaction and relief from boredom. Without it, workers may opt for even less desirable relief behaviors (Fisher, 1987, pp. 11–12). Personal factors include boredom proneness, age, physical capacity, cognitive capacity (Drory, 1982, pp. 149–150) and many more. Of course, fit plays a big role, and while different fit concepts are complex and tend to interact with each other, it can be generalized that a better fit leads to higher job satisfaction, organizational commitment and intrinsic motivation (Kristof-Brown, Zimmerman, & B, 2005, p. 316; Kulik & Oldham, 1987, p. 288). For example, fit and preferences play a role in deciding whether a person finds monotonous jobs boring or enjoyable (Loukidou et al., 2009, p. 9; P. C. Smith, 1955, p. 328).

Next to fit, a special focus lies on intrinsic motivation. Ryan and Deci (2000) stated that "Perhaps no single phenomenon reflects the positive potential of human nature as much as intrinsic motivation", and connections between a lack of intrinsic motivation and workplace/leisure boredom have been established multiple times (Gkorezis & Kastritsi, 2017, p. 105; Shin & Grant, 2019, p. 9; Weissinger, Caldwell, & Bandalos, 1992, p. 323). This is in line with the MAC model, as intrinsic motivation is closely related, if not synonymous, with the meaning component. Thus, understanding intrinsic motivation is helpful in understanding workplace boredom. The job characteristic model by Hackman and Oldham (1976, p. 258) describes job factors influencing employee motivation. These include skill variety, task identity, task significance, autonomy and feedback. In line with this argument, perceived task autonomy itself was shown to be negatively correlated to workplace boredom, and boredom was experienced as especially frustrating when caused by low autonomy (van Hooft & van Hooff, 2018, p. 935).

These insights already carry a number of implications for managers who aim to minimize boredom in order to avoid the negative effects, like counterproductive work behavior and high turnover. For example, increasing any component of the job characteristics model should yield a positive effect on motivation and thus reduce boredom. A special focus should always lie on the fit of personal factors and job/task characteristics. Importantly for this paper, the outlined predictors of workplace boredom provide a framework on how boredom and interest can be manipulated in an experimental setting. At the same time, the insights underline that the manipulation of job factors cannot yield a generalizable manipulation, as fit and personal factors play a central role. This is relevant for the methodology, as it stresses the need to control for variation in what is considered boring.

2.3. Psychological contrast of sensations

While personal and task characteristics do affect the perception of tasks, they are not always the only factors at play. Another important factor can be the context of a task. When conceptualizing work, more often than not, it consists of multiple, sequential tasks (Ilgen & Hollenbeck, 1991, p. 173).

In organizational research, tasks are usually treated as single entities with a start and an end. It is rarely considered whether a task influences another task by contrast. For example, an upcoming unpleasant task might lead to procrastination on the current task, no matter how pleasant it is. Even when tasks are not directly dependent on each other, can still be interdependent.

Assimilation and contrast effects are well-known psychological phenomena that can be applied to a number of areas (Herr, Sherman, & Fazio, 1983, pp. 325–327; Sherif, Taub, & Hovland, 1961). According to the concept, stimuli are judged relative to a reference point, and high differences between stimulus and reference point are perceived stronger than they actually are (contrast effects), while minor differences lead to a convergence (assimilation effects). Examples for this have been found in many different areas, be it soft drinks tasting sweeter when compared to low-sucrose drinks (Riskey, Parducci, & Beauchamp, 1979, pp. 172–173) or candidates in job interviews (Wexley, Yukl, Kovacs, & Sanders, 1972, p. 47). The chosen reference point is often influenced by the most recent comparable experience, so it can be influenced trough temporal order (Brickman et al., 1978, p. 918; Kahneman, Diener, & Schwarz, 1999, p. 15). Crucial for this paper, contrast effects do not only apply to physical stimuli, but also to affects. As Colvin, Diener, Pavot, and Allman (1991, p. 491) observe over multiple studies, "an extremely positive event will not necessarily make bad events seem worse, but an extremely positive event might lower the value of moderately good events". Famously, Brickman et al. (1978, pp. 920–921) found that lottery winners become used to their new pleasures rather quickly. Some more incidental findings already indicate that boredom is also affected by contrast effects. For example, after periods of high activity, workers tend to be more bored when faced with "sharp contrasts" (Fisher, 1993, p. 35). Finally, Wojtowicz, Chater, and Loewenstein (2021, pp. 5–6) opportunity-cost model of boredom proposes that people undergo endowment effects regarding their attention when provided with low reference points, as they expect an ongoing level of stimulation.

Shin and Grant (2019, pp. 4, 23) were able to empirically show contrast effects between intrinsic motivation and performance. In two experiments, a significant relationship was found between intrinsic motivation levels of two consecutive tasks. For performance, they found that high intrinsic motivation will lead to worse performance in an uninteresting follow-up task, but not in an interesting one. The outcome was mediated by boredom. Dora et al. (2021, pp. 10–11) find that smartphone breaks at work lead to subsequently higher boredom. These findings could be viewed as first indication that contrast effects influence the strength of experienced boredom.

The examples show that contrast effects are applicable to different affective states. As intrinsic motivation and boredom are closely related, especially the findings on motivation suggest that cross-task effects may apply to boredom as well. The exploration of contrast effects is the core of the research question and the practical implications. Furthermore, the definition of contrast effects and reference points are helpful for the experimental manipulation.

2.4. Formulation of hypotheses

The portrayed background goes a long way in showing the mental processes of boredom and its different dimensions. It also supplies first looks into the relationship between boredom, concentration and cognitive performance. Applying the concept of contrast effects to boredom and its consequences, multiple hypotheses will be developed.

First findings on contrast effects provide evidence that intrinsic motivation is influenced by contrast effects, and incidental evidence points at their existence for boredom as well. In practice, the experienced boredom of a moderately interesting task should thus be higher when preceded by a highly interesting task and lower when preceded by an uninteresting task, as it pronounces the person-environment fit or lack thereof.

> Hypothesis 1a: The boringness of a task will be higher when it is preceded by a significantly more interesting task.
>
> Hypothesis 1b: The boringness of a task will be lower when it is preceded by a significantly less interesting task.

Drawing from the literature on boredom, there is a close relationship with attention. Low concentration (inattention) can be a feature of boredom; however, it is not synonymous with it. Previously, attention and boredom have been examined as separate constructs (e.g. Hunter & Eastwood, 2018; Wilson et al., 2014), as they will in this paper. While no clear causal evidence of that has been presented yet, it should be expected that people are unable to concentrate when bored and that a relief of boredom leads to better concentration. While attention and boredom are often correlated as a result of overlap in measures (e.g. the attentional component of the MAC model), a boredom manipulation for the same task offers the unique chance to isolate the inattention variable. Later in this paper, the issue of causality will be discussed.

> Hypothesis 2a: When boredom is high, participants' concentration will be lower
>
> Hypothesis 2b: When boredom is low, participants' concentration will be higher

Lastly, the relationship between boredom and performance will be explored. It has been proposed before that boredom leads to generally lower task performance (Cummings, Gao, & Thornburg, 2016, p. 289). As first indication for this research, an early study among truck drivers indicated that boredom is negatively correlated to work effectiveness. This relationship was moderated by physical (e.g. age) and mental capacity (Drory, 1982, pp. 149–150). Just as for concentration, it is unclear whether boredom and cognitive performance are causally correlated to each other. When accounting for inattention, it is expected that boredom itself will decrease the ability and/or willingness to perform.

Hypothesis 3a: When boredom is high, performance in cognitive tasks will be lower

Hypothesis 3b: When boredom is low, performance in cognitive tasks will be higher

3. Methodology

In order to test the hypotheses, a randomized, controlled experiment was performed online, designed to mirror a laboratory experiment. Great care was taken to minimize disturbances to the variables examined while still resembling working conditions as true to life as possible. The experiment was conducted through the online platform Gorilla (www.gorilla.sc), which was validated regarding precision and accuracy of data collection (A. Anwyl-Irvine, Dalmaijer, Hodges, & Evershed, 2021; A. L. Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). The platform hosted the tasks and the questionnaires of the experiment and documented the relevant experimental data. It also registered whether participants used a mobile or desktop device, so potential differences could be controlled for. Among all participants, a giveaway for three Amazon gift cards with a 15€ value was held in order to increase participation willingness. Submitting an email address for the giveaway was not mandatory and the results remained anonymous. As all of the participants were German, the language of the experiment was German as well. Instructions and questionnaires can be found in Appendix A.

3.1. Structure and groups

Participants were divided into three groups. To document the cross-task effects of boredom, the two treatment groups were primed with differing amounts of boredom. Group B-C completed a highly boring task first, followed by a moderately boring, task that was held constant. (Note that boringness is a subjective judgement. However, for better readability, tasks will be referred to as boring and interesting depending on which judgement they aimed for.) Group I-C followed the same structure as B-C, with the variation that the first task was supposed to be highly interesting. In analyses that included both task I and task B combined, they were referred to as task 1. The control group C completed only the constant, moderately interesting task C. This group had natural variety in what activities preceded the experiment and was added to register whether the effect only went in one direction. Following the tasks, every participant filled out a questionnaire, registering boredom, inattention and additional information.

3.2. Tasks

For the boring task B, a qualitative underload approach was chosen. Often, these are highly monotonous and passive tasks. A 2014 paper compared multiple boredom inductions across two experiments to identify the most reliable one (Markey, Chin, Vanepps, & Loewenstein, 2014, pp. 239–24. 243, 245). A digitalized peg turning task, based on Festinger and Carlsmith (Festinger & Carlsmith, 1959, p. 205), lead to the highest self-reported boredom in terms of both intensity and discreteness. Accordingly, it became the recommended induction. For the B-C treatment, this task was adopted. It consisted of eight peg icons, arranged in two rows of four. The participants were instructed to continuously click on the peg that was highlighted, after which it turned 90 degrees clockwise and another peg was randomly highlighted. The task went on for 5 minutes.

Treatment I-C was supposed to receive a task that was deemed interesting by the participants. In similar experiments, videos are often used to induce either boredom or interest in comparison tasks (Markey et al., 2014, p. 240). Especially high-paced videos like clips from action movies, are used to induce interest and increase arousal (Fahlman et al., 2013, p. 78; Hunter & Eastwood, 2018, p. 2486). A more active, similarly reliable interest induction could not be identified, so watching a video was chosen as task I. While video mood inductions can be short-lived (Drody, Ralph, Danckert, & Smilek, 2022, p. 11), they have been shown multiple times to be effective in inducing interest (Drody et al., 2022, p. 9; Hunter & Eastwood, 2018, p. 2488; Merrifield & Danckert, 2014, p. 284). For the experiment, 4 different clips were chosen, from which participants were allowed to pick one that sounded the most interesting to them. After choosing, they were able to change their decision and watch a different clip instead if desired. To increase the meaning component, participants were asked to remember the most important details. It could be argued that different clips could influence the results in different ways, however, there are multiple upsides to providing a choice. First off, it was expected to increase the intensity of interest, as not every person deems the same things as entertaining. Subjectivity could be a limiting factor in inducing interest, which the choice helped to mitigate. Furthermore, participants were given more autonomy when choosing, a factor shown to increase intrinsic motivation and interest (Hackman & Oldham, 1976, p. 258; Westgate & Wilson, 2018, p. 690). Choice was similarly utilized by Fahlman et al. (2013, p. 78), leading to a successful induction and an increase in task engagement. Finally, the questionnaire aimed to control for any disruptive effect that different videos could lead to, as it registered the perceived boredom specifically. Regression analyses with one dummy variable per video also confirmed no significant influence on the outcome measures. Thus, the variability in videos is expected to raise reliability, as it increases autonomy and evens out effects that could be rooted in the other aspects. The four videos were:

- A clip from the American sitcom The Office (Dir. Blitz, 2009), chosen for its high-energy humor and high quantity of jokes,
- a truck chase scene from the action movie Terminator 2 (Dir. Cameron, 1991), chosen for its high-paced action content,
- the trailer for Elvis (Dir. Luhrmann, 2022), chosen for its novelty and the prominent use of editing and music,

Table 1: Overview of the groups and structure

Stage	Group B-C	Group I-C	Group C
1	Boring Task B	Interesting Task I	/
2	Constant Task C	Constant Task C	Constant Task C
3	Questionnaires B, C and G	Questionnaires I, C and G	Questionnaires C and G

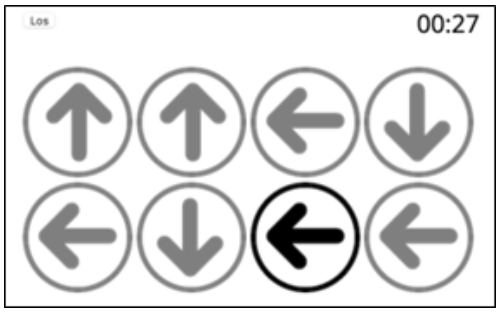


Figure 1: Peg turning task screen

and

- a collection of short videos from the app TikTok (TikTok, n.d.), chosen for their variety and their unexpected character.

The clips were all between 4 and 5 minutes long and chosen to work by themself, i.e. without any context. They were provided in both English and German (except the Tik Tok clip, which was not available in German).

The constant task C, shared among all groups, acted as the main base for comparison between groups. It was chosen to be moderately interesting and challenging while requiring cognitive capacity. For this, a 10x10 Schulte grid, also known as concentration grid, was used. The exercise has been shown to be resistant to practice effects (Greenlees, Thelwell, & Holder, 2006, p. 34). A Schulte grid is a square grid of differing dimensions with equal-sized cells, which is filled with two-digit numbers from 00 to X, in this case 99. Participants were instructed to click on the numbers in ascending order as quick as possible. This exercise requires visual scan speed and higher cognitive capacity will lead to better performance, as the locations of upcoming numbers can be remembered once they are seen. The main variable was the time needed to complete the table. Additionally, errors (i.e. clicking on the wrong number), reminders (i.e. when people forget the next number and click for a reminder) and the score after one minute were registered.

3.3. Questionnaire

After the groups completed their respective tasks, they filled out a questionnaire. This questionnaire had the purpose of registering the participants concentration and boredom along multiple dimensions. Other than the fact that groups B-C and I-C received a questionnaire about two tasks, these only differed in neglectable ways (e.g. when referencing the specific task). The questions on state boredom were drawn from the widely used Multidimensional State Boredom Scale (MSBS). The questions were translated and reworded to refer to the tasks at hand. The scale measures boredom on five subscales and has been tested extensively for validity and reliability ($\alpha = .94$) (Fahlman et al., 2013, pp. 76, 79). As excessively long questionnaires were found to decrease compliance (Sahlqvist et al., 2011, p. 5), the 29 questions in the original MSBS were reduced to 6. Three questions were drawn from the subscale disengagement and one each from high arousal, low arousal and time perception, as these were determined as the ones most applicable to the specific tasks (Fahlman et al., 2013, p. 73). The subscale inattention was separated to assess state concentration individually. Participants were asked whether they had difficulties concentrating, experienced mind wandering (Smallwood & Schooler, 2006, pp. 946–947) were easily distracted during the task. All answers were measured on a 5-point Likert Scale, with 5 indicating the highest boredom. The average of the disengagement, high/low arousal and time perception questions from the MSBS and the question on general boredom were averaged into a boredom score, used as the main variable to measure and process the participants' boredom. The boredom score was treated as interval data, as previously established for Likert scales in similar research (Wu & Leung, 2017, p. 5). Analogously, the three MSBS questions on inattention were combined into the inattention score, where a score of 1 indicated the best concentration and 5 indicates the most difficulties concentrating. Additionally, the groups were asked for their own belief whether they could've scored better in task 2 under different circumstances. Lastly, the

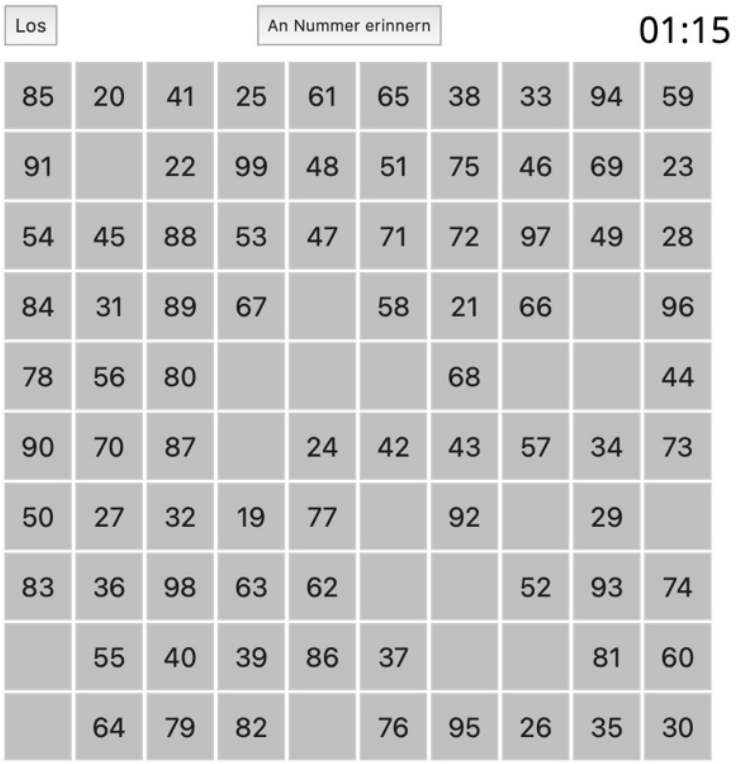


Figure 2: 10x10 Schulte table

candidates answered multiple questions for control purposes, including demographic questions (Age, gender and employment status), three questions from the Boredom Proneness Scale (BPS) to assess their trait boredom (Farmer & Sundberg, 1986, p. 6) and whether they actually watched the video for its full duration. The scale and questionnaires for each task can be found in Appendix A2 - Appendix A6.

3.4. Sample

The experiment was conducted with a representative sample of originally 84 participants. After reduction, discussed in the next section, 75 participants remained. 44 (58.67%) of the participants in the final sample were female, and the average age was 32.06 years (SD = 12.00). With 35 participants (46.67%) of the sample, students made up the highest share of the sample, closely followed by 31 (41.33%) employees. Participants were recruited through social and academic channels, for example by sharing the participation link via social media and in university-related groups. Participants were assigned to the groups fully randomized.

3.5. Validity

Additional measures were taken to attain the highest possible validity. For tasks B and I, the performance was documented in order to assess compliance. Participants who did not actually play the video for the full duration in task I and participants who did not click enough virtual pegs in task B were excluded from the sample. The lower limit for a participant to be included was 260 clicked pegs, below which a gap formed: All of the participants that were excluded clicked less than 100 pegs. Participants who declared in the questionnaire that they did not take part to the best of their abilities were also excluded from the sample before analysis began. Because this research aims to evaluate the cross-task influence of boredom, the groups were reduced by those participants for whom the mood induction was not successful. In group I-C, four participants were removed from consideration for a boredom score above 2.5. Five participants were removed from group B-C, with a boredom score of less than 3.5. 23 participants remained in group B-C, 26 in group I-C and 26 in group C (unchanged), leading to a final sample size of N = 75. As some questions were filled out incorrectly or not at all by some participants, the sample sizes of individual analyses may vary. For example, two participants did not enter their age correctly and were thus excluded from the sample for regressions that considered age as a coefficient.

4. Results

Before reduction, task B received a mean boredom score of 3.94 out of 5 (SD = .75). Task I achieved a mean boredom score of 2.23 (SD = .87), leading to a significant difference between groups, $t(56) = 7.965$, $p < .001$. After reduction, B-C received 4.19 (SD = .53) and I-C 1.94 (SD = .41). The difference between group means after reduction was still highly

significant, $t(47) = 16.595$, $p < .001$. The fact that the difference was significant even before reduction is confirmation that the boredom induction was successful. Other than boredom proneness being negatively correlated to age, $\rho = -.324$, $p < .01$, no correlations among the control variables could be identified.

In the following analyses, a correlation coefficient will be calculated for each pair of outcomes in order to assess the general existence of a relationship. The correlation will be followed by one-sided t-tests to evaluate whether a significant effect exists for both groups and whether the hypothesis can be confirmed. To conclude each analysis, a regression model will be derived to control for other factors and identify strength and significance of treatment effects.

4.1. Cross-task effects on boredom

To explore H1a and H1b, the relevant measurement is the spread in boredom across tasks. If Hypothesis 1a were true, then task C should receive a lower boredom rating in group B-C than it did in the control group C. For group I-C, the score should be higher than that of the control group to confirm H1b. As the second task is the same for everyone task C, significant differences in reported boredom can only be attributed to the context that a task appears in (assuming normal distribution of task-person fit).

To assess whether a correlation between the boredom of both tasks exists independently of groups, Spearman's rho was calculated for the boredom scores of tasks 1 and C. This revealed a moderately strong negative correlation across both groups, $\rho = -.308$. The correlation is significant at $p < .05$. To calculate this correlation, the data for groups B-C and I-C was combined. This combined calculation provides first evidence for the contrast effect hypotheses. The correlation analysis shows that the boredom experienced through a previous task is a significant variable in assessing the boredom of the next task.

The correlation provides evidence for the general existence of cross-task effects; however, it does not consider the symmetry of the effect. In order to assess whether the effect goes in both directions, independent samples t-tests were performed. The tests showed that participants from group B-C ($M = 2.106$, $SD = .905$) experienced significantly less boredom during task C as those from the control group ($M = 2.819$, $SD = .942$), $t(47) = -2.694$, $p < .01$. The same way, they were significantly less bored than those from group I-C, ($M = 2.901$, $SD = 1.07$), $t(47) = -2.783$, $p < .01$. However, no significant difference could be identified between I-C and the control group, $t(50) = .294$, $p = .385$. This speaks to a one-sided effect in which only the high-boredom group B-C provides a significant negative effect on the follow-up tasks boredom.

To explore the relationship further and to control for other potential influences on task boredom, multiple linear regressions were run with Task C boredom as the dependent variable. The tested predictors of a first regression included task 1 boredom as well as boredom proneness, device type and the registered demographic factors (Appendix B1). Overall, the regression was not significant, $R^2 = .274$, $F(8, 37) = 1.800$, $p = .120$. A regression with less factors (Table 2) provided additional insights while being overall significant, $R^2 = .236$, $F(3, 44) = 4.273$, $p < .01$. Part of this regression were device type and gender; however, the only significant coefficient was the boredom score of the preceding task. The regression allows to disentangle the many factors at play and confirms that the significant cross-task effect persists when controlling for personal factors.

Understandably, the task itself remains the biggest influence on boredom. However, the models and tests outlined show a clear cross-task effect of previous boredom. As the pairwise comparison between groups only showed a significance difference for B-C, only Hypotheses 1a can be accepted. This already bears highly interesting insights on how boredom can potentially be reduced through self-organization and managerial intervention.

4.2. Effects on concentration

In order to assess the general relationship between boredom and inattention, Spearman's correlation coefficient was utilized to compare self-reported boredom score and self-reported inattention score. The calculations (Appendix B3) revealed high correlation coefficients of $\rho = .388$ between boredom and inattention for task 1 ($p < .01$) and $\rho = .565$ for task C ($p < .001$). This very high effect is in line with expectations, as boredom and attention are highly correlated constructs and can overlap, as described in chapter 2.1. When looking at the groups individually, the significance is sustained for group I-C. Here, the correlations are $\rho = .515$ for task I ($p < .01$) and $\rho = .737$ for task C ($p < .001$), which is even higher than those for the full sample. For group B-C, only the correlation for task C was significant at $\rho = .489$ ($p < .05$). The correlation for group C is not significant at $\rho = .292$ ($p = .147$). Even though the effects are not significant for every subgroup, the correlations show a clear effect between boredom and concentration, especially for cognitive task C.

Aiming to isolate the direction of the effect, t-tests were performed to compare the means of the task C inattention scores. Overall, group B-C ($M = 1.62$, $SD = .83$) reported the lowest levels of inattention during shared task C, followed shortly by the control group C ($M = 1.85$, $SD = .76$). The group with the highest reported inattention was group I-C ($M = 2.37$, $SD = 1.26$). The t-tests revealed a significant difference between task C inattention levels of I-C and the control group C, $t(50) = 1.778$, $p < .05$, suggesting that the interest induction of task I actively weakened concentration compared to the baseline. Similarly, B-C showed a significant difference from I-C, $t(47) = 2.417$; $p < .01$. However, B-C did not differ significantly from the control group, $t(47) = -1.041$, $p = .15$, meaning that the effect is considered one-sided. The t-tests can thus confirm H2a, but not H2b.

To validate the results and reveal any potential effects of other factors, a multiple linear regression was performed, employing the inattention score of task C as the dependent variable and controlling for demographic factors, device type

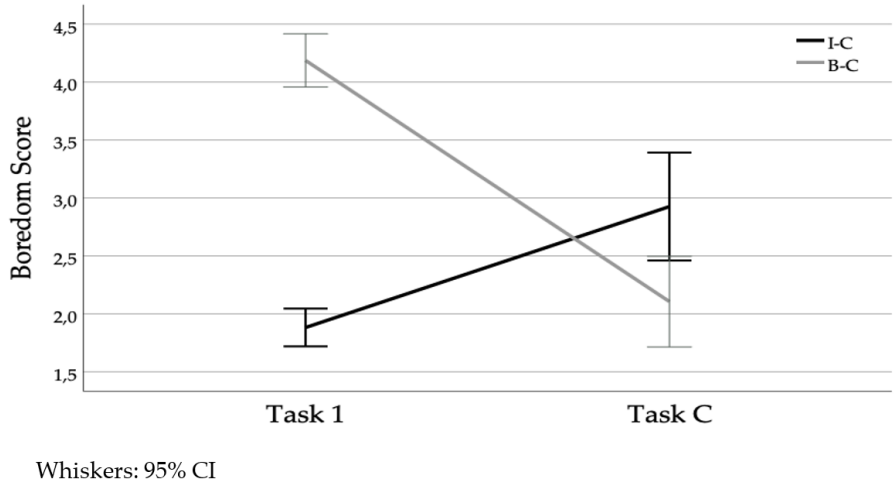


Figure 3: Cross-Task Boredom

Table 2: Regression on sequential Task Boredom

	R^2 (ΔR^2)	B	Std. Error	β	p
Model	.236				.008**
(Constant)		4.677	.610		<.001***
Task 1 Boredom	(.196)	-.417	.124	-.470	.002**
Gender (Female = 1)	(.060)	-.565	.304	-.250	.070
Device (Mobile = 1)	(.066)	-.667	.343	-.271	.058

*p < .05; **p < .01; ***p < .001; N = 48

Table 3: Regression on task C inattention

	R^2 (ΔR^2)	B	Std. Error	β	p
Model	.283				<.001***
(Constant)		.290	.374		.441
Task C Boredom	(.236)	.484	.101	.491	<.001***
Boredom Proneness	(.021)	.139	.097	.147	.155

* p < .05; ** p < .01; *** p < .001; N = 73

and boredom proneness. The first, extensive model was significant, F(8, 62) = 3.095, p < .01 but showed that there are no effects going out from control factors on inattention (Appendix B2). The most significant model (Table 3) included only state boredom and boredom proneness, F(2, 70) = 13.832, p < .001, R^2 = .283. The regression confirms the close relationship between boredom and inattention further and proves that the effects persist when controlling for personal factors. To test whether task C boredom acts as a mediator between task 1 boredom and task C inattention, a mediation analysis was performed. However, as a direct regression of task 1 boredom on task C inattention was not significant, F(1, 47) = 1.226, R^2 = .069, p = .069, a mediator relationship could not be concluded according to the model by Baron and Kenny (1986).

To better assess the causality between boredom and concentration, the inattention scores for task 1 and task C were correlated with each other. This assessment revealed no correlation, ρ = -.007, p = .964. The absence of a correlation indicates that the contrast effects and thus the significant differences between groups are explained by boredom only, as no direct contrast effects exist for concentration. This can be viewed as evidence that inattention was at least partially caused by boredom, as opposed to the other way around.

In conclusion, the exploration of H2 revealed a particularly high correlation between boredom and inattention for most groups, which is in line with expectations and the MAC model. It remains a controversial discussion among researchers whether boredom causes concentration failures or vice versa. While the regression does not allow for a causal

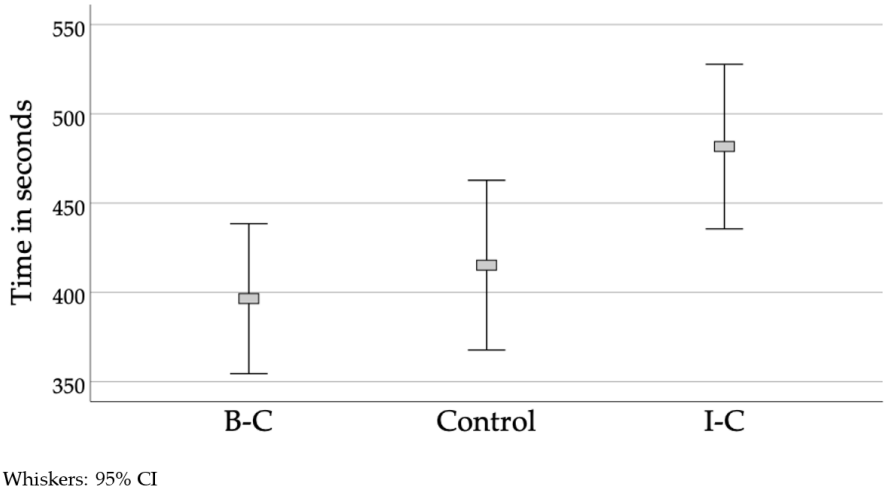


Figure 4: Arithmetic means of Schulte grid times

inference, the absence of contrast effects between inattention scores suggests the former. However, a bidirectional influence is possible and will be discussed later in this paper.

4.3. Effects on cognitive performance

To assess H3, the relationship between self-reported boredom and cognitive performance is observed. If H3a were correct, then group I-C should on average need more time to finish the grid than the control group C did. Analogously, to confirm H3b, the B-C participants should be able to complete it quicker than group C.

Aiming to confirm the internal consistency of the performance measure, correlations were performed among the primary measure (time) and the secondary measures. These showed that the number of errors was positively related to the time needed to complete the Schulte grid, $\rho = .402$, $p < .001$. Similarly, the progress after one minute was negatively correlated to the time needed, $\rho = -.422$, $p < .001$, indicating that the negative effect on performance is homogeneous throughout completion, as opposed to short-lived or late-emerging. Overall, these correlations indicate consistency among the performance measures.

Interestingly, multiple factors were correlated to cognitive performance during the Schulte grid exercise. Both boredom ($\rho = .308$) and inattention ($\rho = .340$) during task C were correlated with the time needed, $p < .01$, which is unsurprising considering the close relationship of the two factors. Additionally, boredom during the first task was by itself significantly correlated with the time needed $\rho = .325$, $p < .05$. Of the secondary performance measures, two were correlated with task C inattention: the achieved Schulte grid score after one minute was positively correlated with them, $\rho = -.228$, $p < .05$, which is consistent with the overall measure. Additionally, inattention correlated to the number of times that participants needed to be reminded of the next number, $\rho = .258$, $p < .05$. These two correlations combined with the higher correlation suggest a close relationship between boredom, inattention and cognitive performance. An overview of correlations can be found in Appendix B4.

Independent samples t-tests were performed for the groups in order to assess the differences in performance (Figure 4). The tests corroborated the results from the correlations, showing a clear significant distinction between the means of group B-C ($M = 396{,}52$, $SD = 97{,}074$) and I-C ($M = 481.65$, $SD = 114{,}079$), $t(47) = 2.794$, $p < .01$. The difference between the two groups was close to 90 seconds. The control group C ($M = 415.27$, $SD = 117{,}663$) did not differ significantly from B-C, $t(47) = .604$, $p = .275$. However, with more than 60 seconds difference, it did differ from I-C, $t(50) = 2.065$, $p < .05$. As group I-C was found to experience more boredom during task C, these results confirm H3a, while H3b cannot be confirmed through the given data.

An explanatory regression was performed in order to control for any additional factors that might influence the results. As the dependent variable, the time to complete the Schulte grid was chosen, as this was the primary performance measure. A first regression (Appendix B5) included task 1 boredom and inattention as well as personal factors and device type. The device type was included because the use of a touchscreen and a smaller display could inhibit the performance. The regression was significant overall, $F(11, 34) = 2.695$, $p < .05$, $R^2 = .466$, but included many insignificant factors. Additionally, due to the inclusion of results from stage 1, the control group was excluded from this regression. The regression showed task C boredom and the device type as the strongest predictors for performance. Another regression model resulted from removing insignificant factors and factors with little to no predictive value (Table 4). As the variables from stage 1 were not included anymore, the regression included the control group again. The model was significant at $p < .001$ and accounts for $R^2 = .315$ of the sample's variance. Again, task C boredom and the device type were the most significant predictors for cognitive performance.

Table 4: Significant regression on Schulte grid time

	R^2 (ΔR^2)	B	Std. Error	β	p
Model	.315				<.001***
(Constant)		172.711	52.602		.002**
Task C Boredom	(.065)	33.359	13.285	.300	.014*
Task C Inattention Score	(.059)	31.848	13.327	.283	.020*
Age	(.051)	2.211	.999	.230	.030*
Dummy: Mobile	(.065)	62.974	25.089	.260	.015*

* $p < .05$; ** $p < .01$; *** $p < .001$; $N = 71$

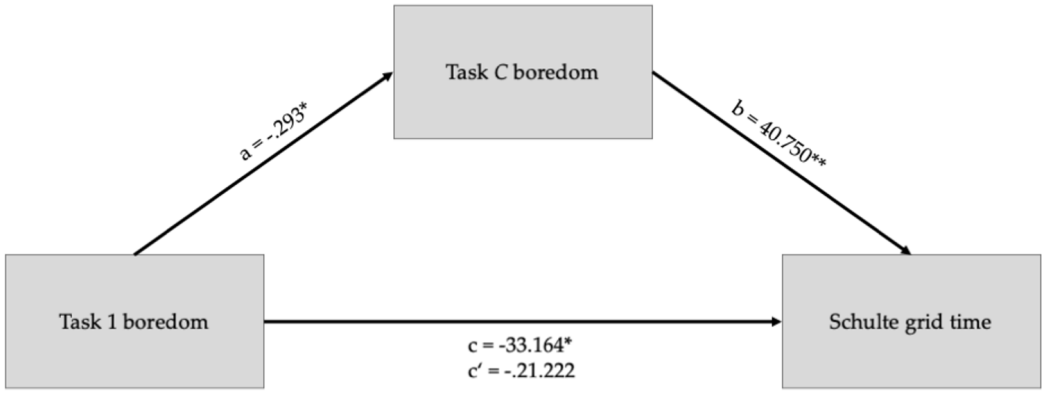


Figure 5: Mediation graph for task boredom and Schulte grid time

The results of the regression suggest that cognitive performance is influenced by age in addition to boredom and inattention. The device type is not interpreted as an influence on cognitive performance, but rather a potential obstruction that was controlled for. As age did not significantly differ between groups B-C (M = 33.43, SD = 13.56), I-C (M = 31.13, SD = 11.55) and Control (M = 31.64, SD = 11.26), age might predominantly explain variance inside the groups, as opposed to variance between the groups. The fact that the boredom score and inattention scores of task C influence the outcome separately in the controlled regression, each yielding a considerable ΔR^2, can be interpreted as further confirmation that boredom and concentration are related, but separate constructs. Importantly, their individual predictive value in the regression model might hint at different separate on performance, for example by leading to unwillingness (boredom) and inability (inattention) to perform.

In order to test whether any variables act as mediators for Schulte grid time, a mediation analysis according to Baron and Kenny (1986) was performed. As the factors task C boredom and inattention did not significantly predict Schulte grid time in a shared regression (Appendix B7), inattention could not be determined as a mediator. However, the mediation analysis yielded one very interesting insight: The effect going out from task 1 boredom on the Schulte table is fully mediated by task C boredom (Appendix B8, Figure 5).

5. Discussion

5.1. Interpretation and limitations of the results

This paper aimed to further the understanding of organizational boredom with a special focus on practical implications. To achieve this, a two-stage experiment was carried out, designed to capture multiple facets of boredom. Specifically, the experiment aimed to understand boredom beyond the scope of just one task and how these cross-task effects influence the outcomes of sequential tasks. In order to test the hypotheses and gain possible explanations for the results, regressions and a questionnaire were utilized. The experiment provided evidence that task boredom is subject to contextual effects and that these effects subsequently influence inattention and cognitive performance.

5.1.1. Contextual effects on boredom levels

In line with expectation and prior findings on psychological contrast effects (see chapter 2.3), the experiment showed significant results on the relativity of boredom in different contexts. Specifically, the findings corroborate the hypothesis that a highly boring first task will mitigate boredom in the subsequent task. These results hold up in comparison to an actively interesting task (I-C) as well as a control group. Correlation and regression analyses confirmed this cross-task relationship. As group I-C does not differ from the control group, this suggests that context effects only arise for a decrease in boredom.

One possible limitation regards the consistency of the effect. As outlined in chapter 2.3, minor differences between stimuli often result in convergence. In this context, that would mean that a task that is only slightly less boring would actually be perceived the same. As minor differences were not examined in this paper, more research is needed to address this possibility and its practical implications. However, the inherently subjective perception of boredom as well as the difficulty of controlling boredom on a miniscule level might limit research possibilities.

More limitations draw from the scales used to assess boredom. Firstly, these scales assessed multiple dimensions, including both low and high arousal. Intuitively, these subscales might impede one another. However, the scale was validated extensively (Fahlman et al., 2013, pp. 75–80) and care was taken to not add questions to the questionnaire that preclude each other. Secondly, the use of 5-point Likert scales could limit the results, as the responses from stage 1 could have acted as reference points for stage 2 and resulted in an anchoring bias and inflated differences. A 7- or 9-point Likert scale might have captured more nuanced results. However, given the validation of the MSBS, it was decided that that the validated 5-point measure should be used throughout the experiment.

Whether the effect is one-sided or not, the study concluded with significant results. Limitations are mainly limited to factors that prevent additional results, and not factors compromising the existing ones. Looking at the significance of results, it can thus be concluded that contrast effects do indeed pertain to state boredom, even though it remains an open question whether this effect is one-sided or not.

5.1.2. Concentration

In addition, it was shown that these contrast effects transcend to boredom's effect on concentration. Among both stages of the experiment, it was thus confirmed that boredom and concentration are highly correlated. This result is unsurprising, as the two constructs are often connected to each other and inattention is even one subscale of the MSBS. Similarly, the MAC model considers an attentional component while still allowing to separate the two constructs. So far, research on the causal relationship between boredom and inattention has tended to focus on sustained attention and vigilance tasks (e.g. Hunter & Eastwood, 2018), but not short-term, practical concentration.

The question whether boredom causes inattention or vice versa remains a controversial one, and existing research has suggested a relationship in both directions (Hunter & Eastwood, 2018, p. 2484). The lack of cross-task correlation between inattention scores in this paper suggests that boredom is the precursor. This does not need to be an either-or question though. It is entirely plausible that both boredom and inattention exhibit bidirectional causality. If this were the case, then boredom would lead to lower concentration, but at the same time low concentration would cause boredom. The relationship between boredom and media multitasking was shown to be bidirectional (Dora et al., 2021, pp. 3, 8), a phenomenon which could be apply to boredom and inattention too. Thus far, research has mainly concentrated on one-directional causal relationships and/or overlap of the two constructs, but a focus on bidirectional causality might lead to interesting insights.

Overall, the experiment concluded a direct, negative relationship between boredom and concentration during a task. Within the experiment, only the relationship between high boredom and low concentration could be confirmed, as group B-C, did not differ significantly from the control group. The natural interpretation of this effect is that the interesting task leads to a persistent distraction afterwards, but the one-sidedness might also come down to limitations of sample size and scales. Overall, the correlations and regressions showed a significant interaction of the two constructs boredom and inattention, with differences between the groups being influenced by the contrast effects outlined earlier in this paper.

5.1.3. Cognitive Performance

Finally, the influences of boredom on cognitive performance were explored. For this, multiple secondary measures (e.g. errors) were taken, showing a conclusive picture and thus confirmed the consistency of the measure. Regressions and comparisons showed a significant difference between groups, indicating that the boredom induction did indeed affect cognitive performance.

Regressions showed an inconclusive picture on the actual effect structure. While the correlation between task C inattention and task performance (measured as time to finish the Schulte grid) was the most significant (Appendix B4), multiple regressions with all factors classified it as fully insignificant and assigned it no explanatory value (Appendix B5). One possible explanation for the inconsistent regression results is the strong collinearity between the factors boredom and inattention in both tasks, which might have distorted the regression (Mason & Perreault, 1991, p. 269). A significant and highly predictive regression model with less factors (Table 4), however, found task C inattention as a significant predictor while controlling for external factors. Independently of the regression, the t-tests confirm the differences between groups.

The regression suggests that both boredom and inattention influence the performance during a task and exert effects of approximately the same strength. Due to the contrast effects, task 1 boredom had an influence on task C performance, which was fully mediated by task C boredom. Task C inattention did not take a mediator role, but still influenced the performance significantly. The two separate predictors boredom and inattention might be interpreted as unwillingness (boredom) and inability (inattention) to engage in the task at hand. Similar to the validated MAC model of boredom, this means that even one of the two factors inattention and boredom is enough to decrease cognitive performance. As additional factors, both age and the type of device (mobile or computer) were identified and included in the regression. Interestingly, the participants in the treatment groups were aware of their performance differences: Groups B-C (M =

2.96, SD = 1.33) and I-C (M = 4.00, SD = 1.06) differed significantly in their belief on whether they could have performed better in task 2, $t(47) = -3.054$, $p < .01$.

5.1.4. Further Insights and Limitations

Throughout the experiment, one interesting insight was the migration of effects from B-C to I-C. While for H1, B-C was significantly different from the other two groups, for H2 and H3, I-C was the one that differed. This effect draws from the position of the control group, while B-C and I-C remain steadily separated. Possible explanations are additional, not registered effects, which would also explain the disproportionately high inattention for group I-C. One possible explanation is simply the time lag between tasks. For group B-C and I-C, the two tasks followed one another right away, while the control group participants were eased into the experiment and had a high variation in what cognitive state they entered the experiment in. Accordingly, group I-C might have been more distracted through the recent video task than the control group. Overall, the validity of the control group can be viewed as a possible limitation. The questionnaire registered the last task completed before the experiment and did not find any imbalances between expected boring or interesting tasks, but the boredom before the beginning of the experiment was not controlled for. The same limitation could be extended to the whole experiment, as it was performed online and could not control for environmental or mood differences as thoroughly as a laboratory experiment.

An additional limitation is that performance could not be compared between task I and task C, as they differed in content. It is suggested that contrast effects are just as strong between leisure and work activities as they are between two work tasks (Dora et al., 2021, p. 11). As the role of the strength of contrast was not assessed, it remains unclear whether a boring first task itself could lead to additional negative consequences. This limitation could have been addressed through a different experimental design where the order between two tasks is changed and the combined performance is compared. The chosen design was opted for instead in order to guarantee a stronger boredom/interest induction and to explore the one-sidedness of the effect through a control group. However, as the control group started with task C, a complete workflow can be approximated as the sum of task C results from groups C + B-C and groups C + I-C. Assessing the performance this way, the results and implications hold up. Furthermore, as one task will always be the most boring, so the negative effects of boredom will emerge no matter when it is performed. Given this, starting with the most boring task will only yield benefits over the span of the subsequent tasks.

The fluctuation of the control group influences the interpretation of the actual effects. While H1 indicated a beneficial effect going out from boredom, the data for H2 and H3 suggests that boredom only neutralizes negative effects, as it does not significantly differ from the control group. It is thinkable that the effect goes in both directions, and the insignificance of effects only draws from the inadequate sample and scale size. It is noticeable that the three groups remain at a constant ranking – the control group is continuously placed in between B-C and I-C (Table 5). For example, the performance difference between B-C and the control group would have been significant at a group size of 250, assuming a one-sided t-test with same mean, standard deviation and a 95% confidence interval.

Next to leading to a more nuanced significance, a higher and more representative sample size could also have helped with regression analysis. For example, only three participants reported their employment status as unemployed, and all three were randomly assigned to group B-C. Accordingly, their effects did not have a base for comparison and limited the regression analysis when controlling for an effect. However, the effects with a sufficiently high sample size (e.g. students) were found to not influence the treatment effects, validating the effects attributed to the treatments.

Overall, even though sample size, chosen scales and the control group exert possible limitations, it can be concluded that the effects between treatment groups are conclusive and consistent. This study adds to our understanding of boredom by providing evidence on the existence of contrast effects and their follow-up effects on concentration and cognitive performance. Beyond the scope of the research question, the interaction between boredom and inattention is now thought to be a bidirectional one, concluding a possible starting point for future research.

5.2. Practical implications

Given the prevalence of boredom in the workplace and the wide consequences that can be caused by it, specifically among young people, this research can provide some practical guidance for work design and managerial practice.

Mainly, the results support the proposed prioritization method for tasks at work. While some tasks are unchangeably boring or interesting to a person, the task order could be a practical way to minimize boredom without side effects. This method suggests that workers begin a workflow with the most boring task and arrange the tasks in increasing order of interest. Because in contrast the tasks will be less boring, this is expected to decrease some negative outcomes associated with boredom. The effects on concentration and cognitive performance have already been proven in this paper, however, it can only be hypothesized that these effects apply to other effects, like counterproductive work behavior, as well. Of course, in practice this prioritization method is limited by other factors like deadlines, but even when considering this, additional benefits can be gained by utilizing the task order to neutralize boredom.

Additionally, the results imply certain concerns for work design. Workers and managers should be aware that smartphone breaks (when deemed more interesting than work) could subsequently increase boredom and decrease performance. Dora et al. (2021) concluded that “smartphone breaks were associated with subsequent increases in, and not with recovery from, boredom and fatigue”, which supports this implication as well. It goes without saying that

Table 5: Comparison of each experimental variable per group

	B-C		Control		I-C
Task C Mean Boredom	2.11	<	2.81	<	2.90
Task C Mean Inattention	1.62	<	1.86	<	2.37
Mean Time Task C	396.52	<	415.27	<	481.65

breaks are a necessary part of work, but it is helpful to be aware of the effects. For example, it might be advisable to either start with a boring task after a break, in order to decrease subsequent boredom. For both options, it is important that the tasks are ordered according to their boringness as much as practically possible. Beyond assessing contrast effects, person-environment fit should always be considered, in order to decrease potential boredom in the first place and enhance job satisfaction in the long run.

5.3. Implications and suggestions for further research

The insights brought up through this paper have a number of implications for boredom research, a field which is highly dynamic at this time. Especially the relationships between boredom and media use or computer-aided work are a recent topic of interest (e.g. Barkley & Lepp, 2021).

One of the longest-lasting and most controversial questions among boredom researchers is the relationship between boredom and attention. As outlined earlier in this paper, researchers are divided on which state is the predictor for the other. In the discussion, the idea of a bidirectional influence was presented, meaning that inattention will increase boredom and vice versa, with both conceivable as the independent variable. Prospective research could dive deeper into this relationship by isolating both boredom and attention systematically.

One way that this paper specifically could be enhanced is through the inclusion of qualitative overload. The experiment focused on qualitative underload, i.e. a task that induced boredom by being underchallenging and monotonous. Interesting insights could be gained by replicating the experiment with an overwhelmingly hard task, which would also induce a feeling of boredom. A possible task could be reading a highly complex science paper, which requires previous knowledge to fully understand. This task would be both challenging and passive, favoring the emergence of boredom.

Finally, replicative studies could aim to reveal a more nuanced picture of the topics discussed. On one hand, this could be done by removing the limitations discussed in chapter 5.1, for example through a higher sample size or a wider Likert scale. But to gain additional knowledge, the levels of boredom could be more nuanced. This way, possible assimilation effects or curvilinear relationships would be revealed. Additionally, the role of the actual strength of contrast could be quantified, for example when the comparison is not between boredom and interest but between high and low boredom/interest. In future iterations of the experiment, it would also be interesting to explore the endurance of the effect. The scope of this experiment rarely exceeded the 10-minute-mark, so no definite estimate could be given of how long-lived the effects actually are. Lastly, field studies would be an interesting addition to the laboratory-style research of this paper. This would bring new insights on the effects the proposed prioritization method in an actual work setting, including long-time performance and compatibility with other prioritization methods.

6. Conclusion

The goal of this paper was to determine potential effects that the task order can exert towards feelings of boredom and subsequently towards concentration and cognitive performance. When exploring the possible consequences of boredom, the role of task order and contrast effects were mostly ignored by researchers. Through an empirical study, this gap in recent research has been partially filled. However, more research is needed to obtain a complete picture and fully understand all relationships.

To answer the research question, an experiment was planned and carried out. The participants were divided into three groups, one of which started with an interesting task and one with a boring task. A control group started with the second stage of the experiment right away. Comparability was ensured by holding the task in the second stage of the experiment constant. Further control measures were taken through a questionnaire, registering both inattention and boredom of the participants during all tasks as well as additional control variables. Multiple measures were compared for each dimension of the research question. In addition, the groups were validated through systematic removal of participants that did not meet the boredom criteria, and multiple regression analyses were utilized to control for other factors that could affect the results.

The results of the experiment confirmed multiple of the hypotheses. Firstly, the existence of contrast effects regarding boredom could be confirmed. This means that more experienced boredom in one task will result in disproportionally less boredom in a following task, given that it was considered less boring in the first place. However, this effect could only be fully confirmed for one direction, namely a decrease in subsequent boredom. Looking at the relationship between high-interest tasks and subsequent boredom, the null hypothesis could not be rejected, even though a significant negative correlation exists for the whole sample. For practice, this carries implications on how workflows can be designed in order to minimize boredom and thus prevent at least some of the

negative outcomes associated with it. The idea for this is to order tasks in order of increasing boringness, so the subsequent task will be considered less boring due to contrast effects.

The outcomes that the effects were explicitly assessed for were concentration and cognitive performance. Concentration was measured as inattention through the questionnaire. For both groups, a very high correlation was found between boredom and inattention. Due to the proximity between the control group and group B-C, the effect could again only be confirmed in one direction. The implication is that inattention arises specifically when a task was preceded by a very interesting, and thus possibly distracting, task. While no additional benefit could be proven for group B-C, the results show that boredom could still be a valuable tool to neutralize any negative consequences.

The findings on cognitive performance were analogue to those on concentration. The boredom group of the sample had a lower arithmetic mean in regard to time to completion, but the difference was not found to be significant. Again, the group that started with an interesting task performed worse than the other two. This is interpreted as a counterproductive effect exerted by the interest induction in stage 1, as the control group acts as the baseline. In addition to boredom, inattention showed an effect on task performance as well. Whether this effect was completely separate or only mediated by concentration could not be clearly determined, with the results from the regression and mediation analysis suggesting the former. Even though the full construct of relationships will have to be explored in future research, the high difference of more than a minute exceeded the expected difference and constitute a big impact in practice, depending on how persistent the effect is.

After concluding the results, possible explanations were discussed and limitations of the experiment were evaluated. While some limitations were found and should be addressed in subsequent research, the overall results were found to be significant and carry many implications for organizational practice and beyond. Boredom was found to negatively affect concentration and performance, and as it was found that the effects of boredom or interest can carry over to the subsequent task of a workflow. Given this, managers and workers should always consider the order of tasks as well as the specific fit with a given activity, not only to maximize performance but also to possibly increase overall job satisfaction and decrease the risk for counter-productive work behavior. Not every task can be easy or enjoyable for everyone, but with the right approach, it might at least become a bit easier and a bit more enjoyable.

References

Agnoli, S., Vanucci, M., Pelagatti, C., & Corazza, G. E. (2018). Exploring the link between mind wandering, mindfulness, and creativity: A multidimensional approach. *Creativity Research Journal*, *30*, 41-53.

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., & Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behavior Research Methods*, *53*, 1407-1425.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388-407.

Barbalet, J. M. (1999). *Boredom and social meaning.*

Barkley, J. E., & Lepp, A. (2021). The effects of smartphone facilitated social media use, treadmill walking, and schoolwork on boredom in college students: Results of a within subjects, controlled experiment. *Computers in Human Behavior*, *114*.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182.

Blitz, J. (2009). *The office (us), s05e14 "stress relief"*. NBC. Retrieved from https://www.youtube.com/watch?v=gO8N3L_aERg

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). *Lottery winners and accident victims: Is happiness relative?* (Vol. 36).

Bruursema, K., Kessler, S. R., & Spector, P. E. (2011). Bored employees misbehaving: The relationship between boredom and counterproductive work behaviour. *Work and Stress*, *25*, 93-107.

Cameron, J. (1991). *Terminator 2: Judgment day.* TriStar Pictures. Retrieved from https://www.youtube.com/watch?v=6z9qws7M8q8

Colvin, C. R., Diener, E., Pavot, W. G., & Allman, A. (1991). The psychic costs of intense positive affect. *Journal of Personality and Social Psychology*, *61*.

Cummings, M. L., Gao, F., & Thornburg, K. M. (2016). Boredom in the workplace: A new look at an old problem. *Human Factors*, *58*, 279-300.

Dora, J., Hooff, M. V., Geurts, S., Kompier, M., & Bijleveld, E. (2021). Fatigue, boredom and objectively measured smartphone use at work. *Royal Society Open Science*, *8*.

Drody, A. C., Ralph, B. C., Danckert, J., & Smilek, D. (2022). Boredom and media multitasking. *Frontiers in Psychology*, *13*.

Drory, A. (1982). Individual differences in boredom proneness and task effectiveness at work. *Personnel Psychology*, *35*, 141-151.

Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science*, *7*, 482-495.

Fahlman, S. A., Mercer-Lynn, K. B., Flora, D. B., & Eastwood, J. D. (2013). Development and validation of the multidimensional state boredom scale. *Assessment*, *20*, 68-85.

Farmer, R., & Sundberg, N. D. (1986). Boredom proneness: The development and correlates of a new scale. *Journal of Personality Assessment*, *50*, 4-17.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203-210.

Firstup.io. (2021). *Key statistics about milennials in the workforce.* Retrieved from https://firstup.io/blog/key-statistics-millennials-in-the-workplace/

Fisher, C. D. (1987). *Boredom: Construct, causes and consequences.*

Fisher, C. D. (1993). Boredom at work: A neglected concept. *Human Relations*, *46*, 395-417.

Gasper, K., & Middlewood, B. L. (2014). Approaching novel thoughts: Understanding why elation and boredom promote associative thought more than distress and relaxation. *Journal of Experimental Social Psychology*, *52*, 50-57.

Gkorezis, P., & Kastritsi, A. (2017). Employee expectations and intrinsic motivation: work-related boredom as a mediator. *Employee Relations*, *39*, 100-111.

Goldberg, Y. K., Eastwood, J. D., LaGuardia, J., & Danckert, J. (2011). Boredom: An emotional experience distinct from apathy, anhedonia, or depression. *Journal of Social and Clinical Psychology*, *30*, 647-666.

Greenlees, I., Thelwell, R., & Holder, T. (2006). Examining the efficacy of the concentration grid exercise as a concentration enhancement exercise. *Psychology of Sport and Exercise*, *7*, 29-39.

Greenson, R. R. (1953). On boredom. *Journal of the American Psychoanalytic Association*, *1*, 7-21.

Haager, J. S., Kuhbandner, C., & Pekrun, R. (2018). To be bored or not to be bored—how task-related boredom influences creative performance. *Journal of Creative Behavior*, *52*, 297-304.

Hackman, J. R., & Oldham, G. R. (1976). Motivation through the design of work: Test of a theory. *Organizational Behavior & Human Performance*, *16*, 250-279.

Hamilton, J. A., Haier, R. J., & Buchsbaum, M. S. (1984). Intrinsic enjoyment and boredom coping scales: Validation with personality, evoked potential and attention measures. *Personality and Individual Differences*, *5*, 183-193.

Harju, L., Hakanen, J. J., & Schaufeli, W. B. (2014). Job boredom and its correlates in 87 finnish organizations. *Journal of Occupational and Environmental Medicine*, *56*, 911-918.

Herr, P. M., Sherman, S. J., & Fazio, R. H. (1983). On the consequences of priming: Assimilation and contrast effects. *Journal of Experimental Social Psychology*, *19*, 323-340.

Hunter, A., & Eastwood, J. D. (2018). Does state boredom cause failures of attention? examining the relations between trait boredom, state boredom, and sustained attention. *Experimental Brain Research*, *236*, 2483-2492.

Ilgen, D. R., & Hollenbeck, J. (1991). The structure of work: Job design and roles. In (p. 165-207). Consulting Psychologists Press.

Johnsen, R. (2016). Boredom and organization studies. *Organization Studies*, *37*, 1403-1415.

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.

Kristof-Brown, A. L., Zimmerman, R. D., & B, E. C. J. H. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology*, *58*, 281-342.

Kulik, C. T., & Oldham, G. R. (1987). *Work design as an approach to person-environment fit* (Vol. 31).

Lipps, T. (1903). *Leitfaden der Psychologie*. Wilhelm Engelmann Verlag.

Loukidou, L., Loan-Clarke, J., & Daniels, K. (2009). Boredom in the workplace: More than monotonous tasks. *International Journal of Management Reviews*, *11*, 381-405.

Lufkin, B. (2021). *The damaging effects of 'boreout' at work.* Retrieved from https://www.bbc.com/worklife/article/20210701-the-damaging-effects-of-boreout-at-work

Luhrmann, B. (2022). *Elvis.* Warner Bros. Retrieved from https://www.youtube.com/watch?time_continue=2&v=KT3GNpxvI00&feature=emb_title

Mael, F., & Jex, S. (2015). Workplace boredom: An integrative model of traditional and contemporary approaches. *Group and Organization Management*, *40*, 131-159.

Mann, S., & Cadman, R. (2014). Does being bored make us more creative? *Creativity Research Journal*, *26*, 165-173.

Markey, A., Chin, A., Vanepps, E. M., & Loewenstein, G. (2014). Identifying a reliable boredom induction. *Perceptual and Motor Skills*, *119*, 237-253.

Martin, M., Sadlo, G., & Stew, G. (2006). The phenomenon of boredom. *Qualitative Research in Psychology*, *3*, 193-211.

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research*, *28*, 268-280.

Merrifield, C., & Danckert, J. (2014). Characterizing the psychophysiological signature of boredom. *Experimental Brain Research*, *232*, 481-491.

Metin, U. B., Taris, T. W., & Peeters, M. C. (2016). Measuring procrastination at work and its associated workplace aspects. *Personality and Individual Differences*, *101*, 254-263.

O'Hanlon, J. F. (1981). *Boredom: Practical consequences and a theory* (Vol. 49).

Preiss, D. D., Cosmelli, D., Grau, V., & Ortiz, D. (2016). Examining the influence of mind wandering and metacognition on creativity in university and vocational students. *Learning and Individual Differences*, *51*, 417-426.

Reijseger, G., Schaufeli, W. B., Peeters, M. C., Taris, T. W., van Beek, I., & Ouweneel, E. (2013). Watching the paint dry at work: psychometric examination of the dutch boredom scale. *Anxiety, Stress and Coping*,

26, 508-525.
Riskey, D. R., Parducci, A., & Beauchamp, G. K. (1979). *Effects of context in judgments of sweetness and pleasantness* (Vol. 26).
Russell, B. (1932). *The conquest of happiness* (5th ed.). George Allen & Unwin Ltd.
Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being self-determination theory. *American Psychologist*, *55*, 68-78.
Sahlqvist, S., Song, Y., Bull, F., Adams, E., Preston, J., & Ogilvie, D. (2011). Effect of questionnaire length, personalisation and reminder type on response rate to a complex postal survey: randomised controlled trial. *BMC Medical Research Methodology*, *11*, 62.
Sherif, M., Taub, D., & Hovland, C. I. (1961). *Social judgment: Assimilation and contrast effects in communication and attitude change.* Yale University Press.
Shin, J., & Grant, A. M. (2019). Bored by interest: Intrinsic motivation in one task can reduce performance on other tasks. *AMJ*, *62*, 415-436.
Smallwood, J., & Schooler, J. W. (2006). The restless mind. *Psychological Bulletin*, *132*, 946-958.
Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*, 813-838.
Smith, P. C. (1955). *The prediction of individual differences in susceptibility to industrial monotony 1* (Vol. 39).
Spector, P. E., Fox, S., Penney, L. M., Bruursema, K., Goh, A., & Kessler, S. (2006). The dimensionality of counterproductivity: Are all counterproductive behaviors created equal? *Journal of Vocational Behavior*, *68*, 446-460.
TikTok. (n.d.). *Tiktok compilation.* ByteDance. Retrieved from https://www.youtube.com/watch?v=VPa_pF5UFi0&feature=emb_imp_woyt
Udemy. (2016). *2016 udemy workplace boredom study.*
van Hooff, M. L., & van Hooft, E. A. (2014). Boredom at work: Proximal and distal consequences of affective work-related boredom. *Journal of Occupational Health Psychology*, *19*, 348-359.
van Hooft, E. A., & van Hooff, M. L. (2018). The state of boredom: Frustrating or depressing? *Motivation and Emotion*, *42*, 931-946.
van Tilburg, W. A., & Igou, E. R. (2017). Boredom begs to differ: Differentiation from other negative emotions. *Emotion*, *17*, 309-322.
Watson, D., Clark, L. A., & Carey, G. (1988). Positive and negative affectivity and their relation to anxiety and depressive disorders. *Journal of Abnormal Psychology*, *97*, 346-333.
Weissinger, E., Caldwell, L. L., & Bandalos, D. L. (1992). Relation between intrinsic motivation and boredom in leisure time. *Leisure Sciences*, *14*, 317-325.
Westgate, E. C., & Steidle, B. (2020). Lost by definition: Why boredom matters for psychology and society. *Social and Personality Psychology Compass*, *14*.
Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The mac model of boredom and cognitive engagement. *Psychological Review*, *125*, 689-713.
Wexley, K. N., Yukl, G. A., Kovacs, S. Z., & Sanders, R. E. (1972). *Importance of contrast effects in employment interviews* (Vol. 56).
White, J. (2014). *Millennial generation eager to work, 'but on their terms'.*
Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., ... Shaked, A. (2014). Just think: The challenges of the disengaged mind. *Science*, *345*, 75-77.
Winograd, M., & Hais, M. (2014). *How millennials could upend Wall Street and corporate America.* Retrieved from http://finance.yahoo.com/news/live-facebooks-investor-call-whatsapp-225719757.
Wojtowicz, Z., Chater, N., & Loewenstein, G. (2021). *Boredom and flow: An opportunity cost theory of attention-directing motivational states.* Retrieved from https://ssrn.com/abstract=3339123
Wu, H., & Leung, S.-O. (2017). Can likert scales be treated as interval scales?—a simulation study. *Journal of Social Service Research*, *43*, 527-532.

Junior Management Science 8(3) (2023) 634-657

Junior Management Science

journal homepage: www.jums.academy

Private Equity Transactions: Value Creation through Operational Engineering Evidence from Europe

Victor Heinrich

Technische Universität München

Abstract

This paper investigates private equity value creation strategies through operational engineering. To examine this, I define a KPI framework typically favored by private equity firms. I apply propensity score matching to a dataset of European PE transactions compared to non-PE backed companies to study value creation. By applying a Difference and Difference regression setting and thereby controlling for two-way fixed effects, I can find strong evidence on PE value creation through operational engineering. This paper adds new insights to academia as (a) there are only few contributions using propensity score matching to examine PE value creation and (b) this paper is the first, to the best of my knowledge, to combine the approach of propensity score matching and Difference in Difference regressions, yielding highly significant results on the relevance of EBITDA margin improvement.

Keywords: Private equity; Value creation; Operational engineering; Propensity score matching.

1. Introduction

> *"People used to think that Private Equity was basically just a compensation scheme, but it is much more about making companies more efficient."*
>
> - David Rubenstein[1]

In the last decade, Private Equity (PE) investments in the European Union grew by more than 10% p.a., setting new records by both deal numbers and transaction volume every year.[2] For instance, in 2020, Thyssenkrupp Elevator AG was acquired by PE investors for 17.2 billion Euros which marks the largest PE transaction on the European market.[3] The market grew particularly strong in Europe as it was barely existent in the 1980s, the first boom-phase of PE in the United States (US). Nevertheless, also in the US, the number of PE transactions has almost doubled between 2000 and 2005.[4] Possible causes of the rapid growth in PE transactions are the anticipation of excess returns and lack of alternatives within the strained capital market due to low interest rates and financial crises, especially for institutional investors.

Before describing the PE market and its characteristics in more detail, one should get an overview of the peculiarities of this asset class. Generally, PE is referred to as the acquisition of equity securities in unlisted companies, which is why PE is considered as an alternative asset class. As this usually entails large transaction volumes, PE is primarily used by institutional investors and wealthy individuals.

While PE funds report record-breaking financials in the last years, this was not always the case. The PE industry appears to be subject to strong cyclical fluctuations. Therefore, one should carefully observe this development as the assumption of PE firms creating excess economic value through their actions has become blurred within the last decades, increasingly questioning the high costs associated with PE investments.[5] In addition to the industry's euphoria in recent decades, a growing number of critical voices in the academic discourse have come up questioning the validity of PE firms' business models. For instance, Guo *et al.* interrogate, whether PE transactions are still capable of creat-

[1] Sender (2013) in Financial Times.

[2] See PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft, 2020, pp. 18-21.

[3] See Knitterscheidt and Murphy (2020).

[4] See Acharya, Franks, and Servaes (2007, p. 1); Sensoy, Wang, and Weisbach (2014, pp. 1-2).

[5] See Braun, Jenkinson, and Schemmerl (2016, p. 1).

DOI: https://doi.org/10.5282/jums/v8i3pp634-657

ing value,[6] Stafford even argues that PE transactions are a scheme for funds to charge high fees as he can replicate their returns with a comparable risk and return pattern using publicly traded securities and homemade leverage.[7] Indeed, the high fees associated with PE investments (mainly attributable to carried interest, management fees, and monitoring and transaction fees) pose additional challenges, as an even higher profit has to be generated to cover these costs.[8]

This paper therefore examines whether PE firms create real economic value by quantifying the operating performance measured by pre-defined key performance indicators (KPI) of 406 leveraged buyouts (LBO) of European based companies between 2013 and 2019. I will also compare these returns with 2,062 transactions from non-PE institutional investors. With this approach, I address the question whether it in fact is PE firms as a "superior form of an organization", as suggested by Kaplan and Strömberg,[9] and the LBO structure that creates surplus value or whether the returns of comparable non-PE backed transaction have a similar KPI development in the years following the transaction. What I am most inquisitive about is whether one can see different pre-buyout characteristics and quantify different development patterns after buyout by matching treatment and control group transactions. This paper therefore contributes to the academic debate on PE value creation in two ways: firstly, by focusing on European-based companies, as previous research predominantly focused on Anglo-Saxon companies and secondly by going beyond the common approach of assessing fund level performance and compare PE-firm to non-PE-backed transactions to assess measures of value creation.

The paper at hand will start with an introduction into the theoretic background of PE as an asset class in chapter 2. Chapter 3 follows with presenting the three main value creation strategies financial, governance, and operational engineering, as defined in academia, and how they can be quantified while also covering critical voices questioning the entire *modus operandi* of PE firms as they might create less value than these firms themselves perceive. After the theoretic framework has been set, chapter 4 will start with developing the research hypotheses to be addressed in this paper. It continues with describing the dataset, the pre-buyout characteristics of the target companies before performing analyses based on propensity score matching (PSM) to compare PE and non-PE transactions. Finally, chapter 5 concludes and discusses the findings, putting them in the framework of academic discourse and giving insights on possible future developments in this industry and avenues for further research.

[6] See Guo, Hotchkiss, and Song (2009, p. 1).
[7] See Stafford (2015, pp. 29-30).
[8] See, for instance, JPMorgan (2021, p. 17).
[9] Kaplan and Strömberg (2009, pp. 130-132).

2. Theoretic Background and Academic Discourse

Between 1990 and 2006, the amounts invested in Private Equity globally has increased fiftyfold and the number of transactions in the US has doubled only between 2000 and 2005 - this untapped growth in number of transactions and thus assets under management appears to have continued steadily in recent years.[10] While having started their first large activities in the 1980s in the US, the PE industry can now be seen as a mature financial sector.[11] This is why it is highly relevant to also approach this topic from an academic perspective. Before chapter 3 covers value creation strategies within the PE industry, this section will address the unique characteristics of PE as alternative asset class. It aims at explaining the asset class itself in chapter 2.1, before section 2.2 will elaborate on the leveraged buyout (LBO), which is the *modus operandi* for most PE transactions. This section closes with a comparison of PE and Venture Capital (VC) as two similar yet distinguishable asset classes within the sphere of alternative investments.

2.1. Private Equity as an Asset Class

PE and alternative investments in general are not uniquely defined. Unlike other alternative asset classes like real estate or currencies, the PE industry is marked by, as the name already suggests, secrecy and often a lack of information on financial figures of companies and transactions. PE firms are usually organized as a limited liability company and act as the general partner (GP) to set up funds which the investors, acting as limited partners (LP), invest in. Usually, PE firms employ highly specialized investment managers and are rather small companies. In fact, PE firms usually are substantially smaller than the companies they target for investments.[12]

Also, within the sphere of PE, one can generally distinguish LBOs and VC as they significantly differ both in what companies are being targeted and how the overall deal financing structure is organized. What is widely referred to as "Private Equity" in academia usually includes LBOs, Growth Capital, and VC.[13] While the transition between the two asset classes is fluent, VC generally refers to investments in less mature private companies. One core idea of VC is to support young and entrepreneurial companies by injecting smaller amounts of equity compared to PE to unleash growth opportunities. This is also why the ticket sizes significantly diverge. Venture capitalists, also alluded to as business angels, bear significantly more risk compared to PE funds since VC usually targets small entrepreneurial companies that do not necessarily have a proven business model or are about to develop it. While the growth potential is huge, so is the

[10] See Acharya et al. (2007, pp. 1-2).
[11] See Puche (2016, p. 5); Sensoy et al. (2014, p. 3).
[12] See Kaplan and Strömberg (2009, p. 123).
[13] See Puche (2016, p. 1).

risk associated to a VC investment. This is also why the expected returns on VC investments of around 40% are significantly higher than for PE investments, with expected returns between 20% and 30%.[14]

Apart from ticket size and the investment's risk and return pattern, the core principle of how these asset classes work, is different. While venture capitalists seek young companies to inject equity for usually a minority stake, PE firms aim at a majority stake or overtaking an entire company with proven and stable business activities while heavily relying on external financing through debt. They do so by employing Leveraged Buyouts as a framework, as extensively displayed in section 2.2. Furthermore, the deal financing structure between PE and VC does significantly differ. While venture capitalists and business angels primarily use equity to invest in target companies, the PE firm's equity stake in LBOs is relatively low.[15] First, PE funds as the GPs raise capital from the committed LPs and secondly use large amounts of debt, which is one characteristic attribute of LBOs.

In line with other publications, I will only include PE-backed LBOs in my definition of PE in this paper.[16] This also has practical reasons, as distinguishing between these two transaction types might be challenging as an identifier it is not included in most commercial databases which ultimately may lead to selection bias.[17]

2.2. PE and its modus operandi: Cyclicality and Buyout Booms

As stated above, PE firms have reached new levels of assets under management. However, this has not always been the case as the entire industry is subject to severe cyclical fluctuations. Acharya *et al.* state that *"(. . .) low interest rate, loose credit conditions and syndication of loans (. . .)"*[18] drive the popularity and amounts of LBOs.

KAPLAN and STRÖMBERG define three major buyout waves in this context: while PE funds first emerged in the 1980s, the first wave lasted for nearly ten years before declining by 1990, again. After this, PE activity significantly increased at the end of this decade, with the second wave peaking in 1998 and finally decreasing with the burst of the dotcom bubble in 2000. The third wave set off in the mid 2000s and reached its climax in 2007 where the PE industry in the US surpassed a valuation of 1% of the US stock market for the first time.[19] Also, due to the low levels of credit spread since 2003, LBOs became even more leveraged and more expensive until the setoff of the financial crisis of 2007/2008.[20]

One reason for this cyclicality is the unique characteristic of PE transactions highly relying on external financing within the framework of an LBO with several special purpose vehicles (SPV). These SPVs are legal entities solely founded as an equity or debt instrument within the transaction. In general, the PE fund as the management company establishes a SPV in the form of a limited liability company[21] and thereby acts as the GP in the investment process. The GP manages the fund and takes all operational decisions while the investor as the LP contributes the equity required besides the debt provided by external credit institutions, which already accounts for 60 to 90% of the buyout price.[22] Besides this classic model, parallel co-investments in a portfolio company through the LP are also possible. This trend became increasingly popular in recent years: while not only the popularity of co-investments grew, some institutional investors also even tend to invest in companies on their own (solo investment) - thereby foregoing the established limited partnership model.[23]

This GP/LP structure has advantages such as the limited liability of the SPVs that are beneficial in the case of negative development of the assets acquired by the fund and impending insolvency. On the other hand, a limited partnership agreement is also associated with high costs for founding and maintaining the SPV ecosystem.

In the framework of a limited partnership, the LPs capital can be drawn whenever the GP has identified a suitable target company. This process of raising capital can be considered as the first phase of the fund lifecycle. Regarding the fund cash flow, the first years of the fund lifecycle where the GP acquires the portfolio companies are marked by negative cash flow because of transaction costs, management fees charged by the GP and maybe even write-offs for failed deals. This period of sourcing the deal flow and targeting firms is the second phase. After the phase of target acquisitions, the fund's third phase of operational improvement takes place ultimately yielding positive cashflows that can be distributed back to the LPs. Finally, the fund lifecycle ends with exiting the investments and divesture. The GP is reimbursed with management fees during the holding period and carried interest while divesture according to contractual agreements that usually include performance hurdles and the committed equity contributions plus capital interest are refunded to the LPs. As one can see, the fund lifecycle starts with negative cashflows and finally yielding positive contributions before ending in divesture. Therefore, one can describe the lifecycle of a PE fund as a so-called "J-curve".

This LBO framework can become rather complex as a single PE firm may use dozens to hundreds of SPVs for a single fund. As they actively engage in every singly portfolio

[14] See Achleitner and Braun (2015, p. 14).
[15] See Achleitner and Braun (2015, pp. 14-15).
[16] See, for instance, Hahn (2009, pp. 12-13).
[17] See Heckman (1979, p. 153).
[18] Acharya, Hahn, and Kehoe (2009, p. 9); See Axelson, Jenkinson, Strömberg, and Weisbach (2012, p. 24); Ljungqvist, Richardson, and Wolfenzon (2008, p. 1).
[19] See Kaplan and Strömberg (2009, pp. 124-127).
[20] See Acharya et al. (2007, p. 3).
[21] For instance, in Germany a GmbH (Gesellschaft mit beschränkter Haftung) or also commonly used the luxembourgish (Société à responsabilité limitée) or dutch (besloten vennootschap met beperkte aansprakelijkheid) equivalents.
[22] See Kaplan and Strömberg (2009, p. 124).
[23] See Fang, Ivashina, and Lerner (2015, p. 160); Braun et al. (2016, pp. 17-18).

company, they try to optimize the business to generate surplus value for the fund and its investors. Therefore, most research in this area has focused on the performance of single funds as this represents the return to the investor.[24] This, however, might not be the best approach. As researchers are often interested in whether (and if yes: how) value is generated by the GP, one should carefully investigate the actions taken by the GP to assess their financial impact on exit valuation. To do this, one needs to focus on deal level data of single transactions. This, however, is even more challenging than evaluating fund level data as it this proprietary information is kept highly secret by the GPs. While most research has covered fund performance, there are only few academic contributions focusing on deal level data to shed light on the value creation process.[25] For this reason, the next chapter will cover value creation strategies applied by PE funds and present the debate in academia.

3. Value Creation in Leveraged Buyouts

The holistic idea of value creation through different levers is at the core of PE fund managers value proposition towards investors. Therefore, it is crucial to methodically understand the value truly generated by fund managers as well as the strategies applied to generate these returns to critically evaluate the risk and return profile associated with alternative and, especially, PE investments. However, the academic discourse on PE value creation is still in an early stage. For this reason, there is no universal tool or generally applicable methodology for measuring overall value creation. With the emergence of PE as an asset class in recent decades, academic interest for this industry and its value creation mechanisms also evolved. In addition to studies investigating fund level performance, few studies on value creation on a deal level arose, though mainly focusing on the US as the largest PE market.[26] Assessing fund performance, however, is not the best suited approach when evaluating the GP's skill on a transactional level as it does not address the question of by which means value verily is generated within a portfolio company that ultimately translates into the fund performance.

Within this discourse of whether to regard performance of a fund or a single portfolio company to assess value creation, some might also argue there is no genuine value creation through LBOs, but only wealth transferred to the GP (value transfer hypothesis).[27] On the other hand, authors also argue that PE ownership does not create any new value, but organizational improvements may lead to increased financial benefits for the LBO stakeholders (value transfer hypothesis). Most empirical studies, however, do confirm that buyout transactions are subject to significant gains in economic value due to an increase in profitability as well as productivity. Other studies, however, also confirm this value creation hypothesis. As can be seen, the academic discourse on this debate is inexhaustive and a range of vantage points and opinions on the PE value creation process have emerged.

Nonetheless, in most cases value increases are caused by a mixture of both value transfer and value creation which is difficult to disentangle into its distinct underlying value drivers. Therefore, Achleitner et al. (2010) pioneered in this area of research and developed a comprehensive framework for methodologically capturing and decomposing the economic value created within a PE transaction using a deal level data set from European buyout transactions: by unlevering returns, one can decompose PE deal returns into their sub-parts, which the authors refer to as the *Value Creation Bridge*. From this, three overall strategies could be employed to achieve increased value: financial, operational, and governance engineering, as displayed in Figure 1.[28]

Besides the levers identified by Achleitner *et al.*, value capturing refers to an increase in value without any changes in financial performance and may occur due to negotiation skill at the time of buyout and divesture. However, the two primary and one secondary levers displayed in Figure 1 do have a direct bottom line effect and lead to direct value creation through the actions taken and implemented.[29] The differentiation into these three main value creation pathways also is widely accepted in academia[30] and can be split into distinctive drivers for each value creation pathway. While financial engineering covers factors mainly implied by the leverage effect, operational engineering focusses on actual improvements due to operative and strategic advice as well as actions imposed by the GP. Thus, the latter effect, namely increasing EBITDA[31] and free cash flow (FCF), does require skill and specialized expertise by the GP. The highly relevant EBITDA effect can also be further decomposed into effects resulting from increased sales and margins. Within the operational engineering framework, also multiple and combination effects are considered. However, as the value creation bridge aims at mathematically decomposing returns, it neglects potential effects of interdependencies and other, unobserved effects. Therefore, governance engineering might be adopted as an overarching strategy. This value creation driver refers to the impact of expanded and optimized monitoring and governing mechanisms within the portfolio company to decrease agency costs.[32]

Given these interdependencies between value creation drivers and the current academic discourse, the next chap-

[24] See Achleitner, Braun, Engel, Figge, and Tappeiner (2010, p. 17).

[25] See Achleitner et al. (2010, pp. 17-18); Acharya et al. (2009, pp. 7-9).

[26] See Achleitner et al. (2010, p. 1, p. 17); Kaplan (1989a, pp. 218-219); Guo et al. (2009, pp. 1-2); Cohn et al. (2020), p. 258; Kaplan and Schoar (2005, pp. 1791-1792); Phalippou and Zollo (2005, p. 2).

[27] See Lowenstein (1985, p. 731).

[28] See Achleitner et al. (2010, p. 18-19).

[29] See Berg and Gottschlag (2003, p. 7).

[30] See Gompers, Mukharlyamov, and Mukharlyamov (2015, pp. 2-3); Kaplan and Strömberg (2009, pp. 130-132); Berg and Gottschlag (2003, pp. 4-9).

[31] EBITDA - Earnings before Interest, Taxes, Depreciation and Amortization

[32] See Biesinger, Bircan, and Ljungqvist (2020, pp. 8-9); Gompers et al. (2016) p. 3.

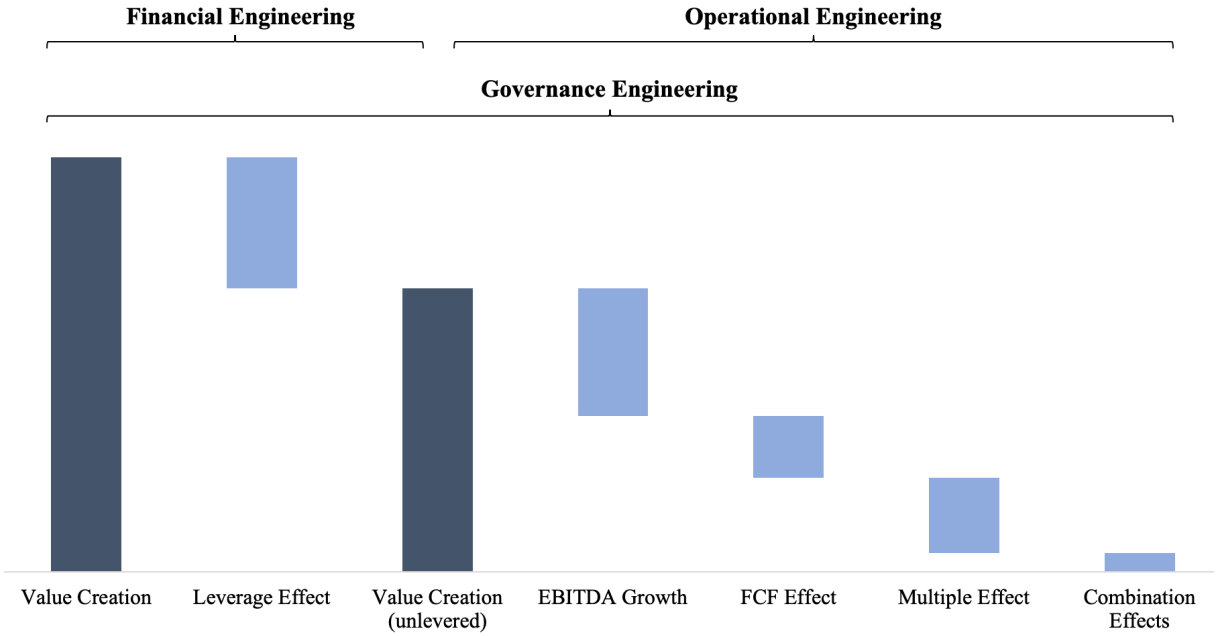


Figure 1: Value Creation Bridge according to Achleitner *et al.*

Own representation based on Achleitner et al. (2010, p. 19).

ters will present strategies and academic findings for each of the three major value creation pathways.

3.1. Financial Engineering

Financial engineering is one unique attribute of LBOs as it refers to the value creation through external financing. Value creation through financial engineering can best be thought of as a value shift enabled by altering the capital structure, primarily via the use of debt, on the portfolio company's balance sheet.[33] The relevance of financial engineering is even stronger in times of low interest rates, as could have been seen in recent years: *"when credit is abundant and cheap, buy-outs become more leveraged"*[34]. By using leverage, companies can lower their capital costs in terms of weighted cost of capital (WACC) and are able to maximize the valuation multiple at the time of exit. Also, decreasing WACC by taking on more debt is cheaper than the costs of the investor's equity, yielding excess returns, and therefore posing the predominant form of deal financing in LBOs.[35]

Besides this, increased debt allows firms to leverage their operating earnings and make use of tax-shield effects to increase the return on invested capital. By reducing taxable income from higher interest and depreciation deductions due to debt repayments, substantial additional value can be generated that can be quantified and translated into higher exit valuation multiples.[36] This effect can even be reinforced when the GP is capable of perceiving market inefficiencies and arbitrage opportunities unveil: Engel, Braun, and Achleitner (2012) show that PE firms in fact have access to underpriced debt for buying equity and profit from this by capitalizing potential market inadequacies between debt and equity market and exploiting their superior information about the target company.[37] While there are academic contributions showing high levels of leverage at buyout with continuous de-levering throughout the holding period, suggesting PE firms in fact exploiting these market inefficiencies,[38] other papers find no clear evidence on leverage development patterns throughout the observation period.[39] Some authors even argue that this exploitation of market inefficiencies set the start for the ever-growing buyout boom of the early 2000s that finally collapsed or even led to the credit market turmoil and the financial crisis starting in 2007.[40]

In turn, increased leverage goes along with a higher risk, which is why an investor as a *homo oeconomicus* may demand a risk premium for higher leverage in a buyout scenario.[41] Additionally, PE is a rather illiquid asset class. Compared to, for instance, publicly traded securities, an investor would therefore, *ceteris paribus*, demand an illiquidity premium for an investment in this asset class.[42] This is also why recent academic contributions usually define financial engineering

[33] See Berg and Gottschlag (2003, p. 7, p. 19).
[34] Axelson et al. (2012, p. 32).
[35] See Guo et al. (2009, p. 27).
[36] See Kaplan (1989b, p. 630-631); Lowenstein (1985, p. 759).
[37] See Engel et al. (2012, p. 487); Berg and Gottschlag (2003, pp. 14-16).
[38] See Achleitner, Braun, and Engel (2011, p. 5).
[39] See Stafford (2015, p. 11, p. 16).
[40] See Kaplan and Strömberg (2009, p. 122).
[41] See Achleitner et al. (2011, p. 3).
[42] See, for instance, Stafford (2015, p. 25-26); Harris, Jenkinson, and Kaplan (2013, p. 4).

as one major driver in PE value creation yielding excess returns. In line with this, leverage effect is supposed to account, depending on the sample and methodology applied, for around 20% to 30% of overall economic value generated in the transaction.[43] However, some studies also find evidence that GPs tend to enter overpriced agreements the more leverage they can use, resulting in lower returns.[44] I can generally distinguish two main results of highly levering a target company: on the one hand, an already profitable company might suffer from financing constraints and therefore lacks capacity to carry out promising and net present value positive investment opportunities. With the capital injected through an LBO, the target company would be able to relax these financing constraints and unleash its growth potential. On the other hand, PE firms might also invest in struggling firms and use the equity to recover the business and capitalize on the business model.[45]

Besides these empirically profound findings just discussed, value creation cannot solely be explained with financial engineering.[46] Also, the relative importance of financial engineering seems to have declined in recent years: while transactions in the early stage of PE activity in the 1980s heavily relied on leverage and governance mechanisms as source of excess returns, portfolio company processes nowadays are typically optimized and enhanced through operational improvements.[47]

It therefore is important to identify additional drivers associated with unlevered private equity value creation. In particular, the origination of the excess returns, which is the difference between the unlevered return of the portfolio company and the unlevered returns of a suitable reference group, i.e., similar companies or industry returns, is relevant for further investigation.[48]

3.2. Governance Engineering

The value creation bridge introduced by Achleitner *et al.* tries to disentangle monetary returns in a quantitative way. However, due to the nature of this concept, it neglects other perspectives that might be worth considering as it might unfold effects across different sections of the framework. As there might be an overlap in the sources of value creation,[49] governance engineering can best be described as an overarching layer addressing all residual value drivers and business processes within an LBO, while not having a direct bottom-line impact.

This value creation strategy refers to effects of increased and optimized supervision and governance mechanisms, the so-called secondary layers, within the portfolio company to reduce agency costs.[50] These secondary levers do not necessarily have a direct bottom line impact but increase it through interactions with the primary levers.[51] Like in publicly traded firms, where executives often receive (virtual) shares and stock options as part of their compensation, this may align interests of different stakeholders and ultimately reduces agency costs. Also, it is quite common to replace the entire management team after a buyout which is one example to what extent PE firms are involved in the newly acquired portfolio company.[52]

Furthermore, standardized planning and monitoring can result in increasing sales volume and profitability, which ultimately creates surplus value.[53] One reason why these governing mechanisms are implemented more successfully in portfolio companies than, for instance, in family businesses is the GP's expertise in the PE sphere or a specific industry. One possible explanation for this are the academic findings that PE excess returns usually are time persistent.[54] This means that the GPs who successfully implement these overarching secondary layers can reduce agency costs within cash flow relevant processes. Apparently, not every PE fund can do so which is why the successful funds seem to have skill rather than pure luck as they can show excess returns continuously across vintage years.[55]

3.3. Operational Engineering

While financial engineering was the primary source of value creation in PE's *"early stage"*, the late 1980s, both practitioners and researchers nowadays mainly focus on actual measures imposed within the portfolio company by the GP. Through these actions, commonly referred to as operational engineering, the PE firm intervenes in the business processes and strategically optimizes them.[56] In fact, most modern and successful PE firms do focus on certain industries which leads to access to industry experts and special knowledge through the GP's network which reinforces the relevance of operational engineering.[57] Today, operational engineering due to operative and strategic advice and improvements actions imposed by the PE firm is the prevailing value creation strategy,[58] in some cases even resulting in abnormal performance[59].

ACHLEITNER *et al.* find that the shift towards operational engineering is even stronger on the European market and accounts for almost half the value created. They identify two

[43] See Achleitner et al. (2010, p. 25).
[44] See Axelson et al. (2012, p. 1).
[45] See Cohn, Hotchkiss, and Towery (2022, pp. 270-271).
[46] See Achleitner et al. (2010, pp. 17-19).
[47] See Kaplan and Strömberg (2009, pp. 132-133); Puche (2016, p. 41).
[48] See Acharya et al. (2009, pp. 14-15); Puche (2016, p. 20).
[49] See, for instance, Guo et al. (2009, p. 3).
[50] See Biesinger et al. (2020, pp. 14-16); Gompers et al. (2015, p. 5); Berg and Gottschlag (2003, pp. 24-30).
[51] See Berg and Gottschlag (2003, p. 24).
[52] See Anders (1992, pp. 8-12).
[53] See Biesinger et al. (2020, pp. 1-2).
[54] Acharya et al. (2009, pp. 11-23).
[55] See Berg and Gottschlag (2003, p. 17, p. 29); Johan and Zhang (2021, p. 217); Jensen (1986, pp. 328-329).
[56] See Graf, Kaserer, and Schmidt (2009, p. 15).
[57] See Kaplan and Strömberg (2009, p. 135); Graf et al. (2009, p. 15).
[58] See Achleitner et al. (2010, pp. 25-26); Harris et al. (2013, p. 20); Kaplan and Strömberg (2009, pp. 131-132).
[59] Achleitner et al. (2010, pp. 25-26); Acharya et al. (2009, pp. 23-24).

major effects within the value creation bridge to capture operational effects: EBITDA growth and the FCF effect, where the latter is mainly affected by working capital optimization, investments, tax, debt (re-)payments and EBIDTDA growth. Also, the excess multiple expansion (defined as the multiple effect that incorporates the change in valuation multiple between entry and exit) represents a fundamental factor in explaining equity returns through operational engineering as a result of a PE fund manager's skill rather than pure luck or macroeconomic fixed effects. As both EBITDA and multiples do have an impact on enterprise value (EV), a correcting factor is added to eliminate effects stemming from the aforementioned value drivers.[60]

Other authors argue that a more efficient usage of existing assets and excising unproductive ones requires skill and therefore is one major underlying sources of operational effects.[61] This reinforces the argument of skill: through efficient cost-cutting measures and strategic decisions taken by experts, value can be generated that exceeds the benefits created through financial engineering. However, literature also presents ambiguous results on operational engineering. While some studies find little to no evidence for operational improvements, most authors do find evidence for it, especially in Europe.[62]

Several authors find significant evidence for increases in operating performance during the first buyout wave in the US in the late 1980s. Kaplan found that his sample of public-to-private transactions systematically outperformed the market through EBITDA growth.[63] Other authors also report findings that are in line with this.[64] On the other hand, more recent studies show a blurred picture: while Acharya *et al.* report significant increases in EBITDA and sales growth, Guo *et al.* see a negative trend after buyout.[65] In summary, operational engineering drivers have had a significant impact on value creation throughout the first buyout wave in the late 1980s. While this value driver appeared to become more relevant also in the 2000s, most funds adapted this strategy and shifted their focus from value creation through excess leverage to value creation through operational improvements within the portfolio companies and the optimization of governance mechanisms.

However, operational effects do not always seem to be clearly significant within the process of genuine value creation. As the relative importance of financial engineering has decreased, governance mechanisms must have gained in relative importance. Alternatively, other KPIs are being improved by "modern" GPs so that the older approaches to capture this value creation cannot account for them. The value creation bridge introduced by Achleitner *et al.* may nevertheless offer a powerful tool to do so.

[60] See Achleitner et al. (2010, pp. 18-19).
[61] See Guo et al. (2009, p. 2).
[62] See Acharya et al. (2009, p. 12); Guo et al. (2009, p. 17); Achleitner et al. (2010, pp. 20-23); Achleitner et al. (2011, pp. 14-25).
[63] See Kaplan (1989a, pp. 250-251).
[64] See Harris et al. (2013, p. 27).
[65] See Guo et al. (2009, p. 28); Acharya et al. (2009, p. 25).

3.4. Why Private Equity Performance is also critically reviewed

Besides the authors who are decomposing value creation and certifying value generation through different levers, one can also observe that the performance persistence of PE as a "superior asset class" is not as clear today as it was in its evolving phase during the 1980s.[66] Results of more recent contributions have shifted the clarity of results - some authors even take a completely different point of view.

LOWENSTEIN, for instance, introduced the concept of the value transfer hypothesis, stating that through an LBO no new value is generated but only transferred to the GP.[67] As discussed above, recent studies find mixed results in terms of value creation especially since the second buyout wave's setoff. Some papers suggest that there is little value creation in PE-backed transactions,[68] other authors even go a step further: STAFFORD showed quite impressively that it would be possible to replicate a portfolio with the same risk and return pattern as a PE fund using homemade leverage and hold-to-maturity accounting. This portfolio with publicly traded securities in fact outperformed PE returns, even before fees, leading to the conclusion that PE does not create surplus economic value and PE investors either take way more risk on than they realize or have severe internal agency conflicts leading to inefficient asset allocation.[69] In line with this, concerns also may arise from return smoothing policies applied by PE firms. Given the assets illiquid nature and that they are not publicly traded, the GP alone values the portfolio company. This allows the PE firm to understate the factual market exposure and thereby artificially downsize portfolio volatility. While this practice can be observed in (hedge) funds, it is very likely to be even more present in private transactions given the industry's secrecy and lack of public reporting requirements.[70]

Therefore, one may wonder why PE is once again on the rise, given investing in this asset class is associated with high costs and uncertainty as well as long holding periods and thus illiquid in nature. Is it only the need for diversification in a low-interest rate environment that drives demand for PE investments? Generally, PE activity rises when interest rates are low as it loosens the credit limits and allows to leverage a portfolio company even more.[71] Also, other asset classes like real estate become more expensive the lower the interest rates are, which may explain the buyout waves of the 1980s and 2000s. However, this trend does not explain whether value is generated through these transactions.

[66] See Braun et al. (2016, p. 1).
[67] See Lowenstein (1985, p. 731).
[68] See Guo et al. (2009, p. 1).
[69] See Stafford (2015, pp. 2-5, p. 28).
[70] See Asness, Krail, and Liew (2001, p. 13); Stafford (2015, p. 4).
[71] See Axelson, Jenkinson, Weisbach, and Strömberg (2008, p. 18, pp. 22-23).

4. Data Analysis

As seen, contrary findings on value creation strategies in LBOs exist. Especially, as the practical relevance of leverage has declined since the 1980s, recent literature finds mixed results on operational effects. Thus, research in the sphere of value creation through operational engineering seems worthwhile to follow. Also, academia appears to lack deal-level information on European transactions, as most literature focusses on fund-level data in the Anglo-Saxon area. Other authors even critically challenge the entire concept of PE investments by simply replicating their returns with a comparable risk and return pattern foregoing the classic GP/LP partnership structure. From this discourse I want to derive the following research question:

> *Are PE-backed transactions in Europe more heavily influenced by operational engineering value creation strategies than non-PE-backed transactions?*

To address the research question, I will formulate three hypotheses that will be tested in this chapter. In this context, section 4.1. develop the hypotheses by motivating and justifying them, before describing the data set and its properties. Chapter 4.3. will give a descriptive overview of the data and will compare treatment and control group characteristics. Section 4.4. will elaborate on the research design and present the methods applied before chapter 4.5. will finally outline the results and findings.

4.1. Hypothesis Derivation

The main question at hand when assessing PE performance nowadays is whether PE funds genuinely create economic value through their actions. I would therefore expect to find significant differences in financial characteristics when comparing LBOs and non-PE backed transactions. Furthermore, given PE firm's intensive commercial and operational due diligence efforts, it also is conceivable that control group transaction and PE firm targets' financial characteristics differ pre buyout. I therefore formulate the first research hypothesis as follows:

> H_1: The KPIs of PE target firms and control group transactions differ significantly pre-buyout.

To evaluate H_1, several metrics might be relevant. With the separation of value creation into financial and operational engineering as suggested by Achleitner *et al.*, amongst others, it seems reasonable to take the operational factors into closer consideration as these are drivers, namely improvements in EBITDA and FCF, are influenced by the GP's action during the holding period.[72]

However, leverage might still have a non-neglectable effect on value creation. For this reason, this KPI will be taken into initial consideration, too. To get an estimate of the firm size, also metrics for size and profitability are relevant. For this reason, I consider logarithmized assets (lnAssets) and sales as a size approximator and EBITDA/sales and FCF/sales margin as profitability parameters. In terms of profitability, companies with overall low levels of profitability might less likely be targeted by a PE investor. On the other hand, comparably unprofitable firms could offer more potential for operational improvements and thereby offer opportunities for value creation. On the contrary, a firm with above average profitability might also not be the desirable target company as it becomes increasingly challenging to capitalize on market momentum and participate in future sales and profitability growth.[73] Thus, I will evaluate the selection pattern used by PE firms with this set of KPIs. Focusing on these KPIs also is judicious for other reasons: first, EBITDA is suitable as a measure for comparing a company's performance. Unlike net income, EBITDA it is not distorted by interest, tax, depreciation, and amortization and thus depicts a company's operational earning capabilities.[74] Therefore, EBITDA can be used to assess a firm's ability to repay debt, a very important information in an mergers and acquisitions setting with highly levered transactions. On the other hand, FCF might be more suitable to assess a company's real valuation, as it is unencumbered. Also, increases in FCF are driven by decreasing capital expenditures (CapEx) and increasing operating income, which captures the potential effects of operational engineering well.[75]

Besides the comparison of pre-buyout characteristics, it is detrimental to observe their development throughout the observation period. If H_1 was to hold true, PE firms would make use of a specific target selection pattern to ultimately generate excess returns through operational engineering. I would therefore expect the PE-backed companies to evolve differently throughout the observation period in terms of EBITDA and FCF as well as profitability than the control group transactions as these KPIs can be perceived as the main drivers of value creation through operational engineering. Following this, I should be able find significantly different KPI developments at a defined level of certainty. I therefore formulate hypothesis two as follows:

> H_2: PE firms do have a target selection pattern based on a set of KPIs that is different to non-PE firms. These KPIs evolve disparately throughout the observation period.

Given the relevance of operational engineering in European transactions, EBITDA and FCF and their post-buyout development are the relevant factors for further evaluation. To account for size-fixed effects, also their sales margins are to be considered.[76] Given the skill and knowledge PE firms apply to create excess economic value through operational engineering, deals backed by PE firms should outperform non-PE

[72] See Achleitner et al. (2010, pp. 18-19).

[73] See Acharya et al. (2009, p. 18).

[74] See Acharya et al. (2009, p. 13).

[75] See Jensen (1986, p. 323, pp. 327-328).

[76] See Cohn et al. (2022, pp. 274-275).

transactions in terms of the above-mentioned KPIs. I therefore formulate my third and last hypothesis:

> H_3: There is a stronger growth in profitability and KPI improvements within the PE-backed treatment group than in the control group. This increase is attributable to operational engineering measures.

These hypotheses will be addressed in chapters 4.3. to 4.5, after the data set and its characteristics have been introduced in the next chapter.

4.2. Data Collection

As there is abundant literature on the deal level data sets in the US, the aim of this paper is to examine performance on the transactional level through operational engineering in European transactions, the second-largest market for PE investments after the US. However, it is not trivial to collect financial data covering deal level PE transactions as the target companies usually do not have to publicly disclose their balance sheets and financial reports and the PE firms being utterly secretive. This complicates retrieving reliable, correct and up to date financial data.[77]

For this reason, I collected two data sets from Bureau van Dijk's Orbis database based on balance sheet and cash flow statement information for each financial year available. With these datasets I can analyze deal-level data as I can calculate KPI developments on a company-level from single financial statements line items (FSLI). I collected two datasets for comparison and analysis: the first contains deal-level data on PE firm-backed LBOs, which I will refer to as the treatment group. The second data set, the control group, contains financial data on non-PE backed transactions. My main sample contains transactions from Austria, France, Germany, Great Britain, Italy, and Switzerland - as the largest economies in Europe and the European G7 countries, amended by Austria and Switzerland for the geographic German speaking GAS region. The final sample includes transactions closed between 2013 and 2019 as this period is in line with the data availability in Orbis. Following Kaplan *et al.* and Guo *et al.*, I will focus on a timeframe before and after the buyout: The year before the buyout (T-1) until two years after the buyout (T+2).[78]

After collecting the data, I manually performed some initial tidying activities before importing the datasets to R Studio.[79] The final sample only includes transactions for which I can calculate all KPIs necessary for further analysis (EBITDA, FCF, assets, and leverage) for the entire observation period. As commercial databases regularly contain self-reported or estimated numbers,[80] I will also only include officially reported financial statements. I excluded non plausible entries such as negative values for sales and converted FSLIs in other currencies into Euro given the year-end exchange rates reported by the European Commission.[81]

After the data is cleaned, I calculate the relevant KPIs from the balance sheets and profit and loss statements for further analysis. To follow the concept introduced by Achleitner et al. (2011) and other authors, I will mainly focus on EBITDA and FCF as KPIs influenced by operational engineering. As the sample consists of deals from different countries, reporting standards and therefore KPIs reported by a company may not always be comparable. Also, neither EBITDA nor FCF are uniquely defined according to generally accepted accounting principles (GAAP) or international financial reporting standards (IFRS),[82] which is why I will use Orbis' KPI definition and calculate the KPIs from the relevant FSLIs.[83] All KPI calculations and definitions used within the course of the next chapters are decomposed in *Appendix 1*.

After I have calculated the KPIs, I added dummy variables for treatment status (treatment vs. control group), buyout year, target country and industry. Overall, Orbis includes 25 default industry classifications. For reasons of simplicity and to avoid potential overfitting of the regression models to follow due to too many dummy variables, I synopsize these subindustries according to a five-industry classification based on the framework introduced by *Fama and French*.[84] The assignment of SIC codes to the five industry types is displayed in Appendix 4. These industries are:

- FF1 Consumer durables (wholesale, retail etc.)
- FF2 Manufacturing, energy, and utilities
- FF3 High-tech, business equipment, telephone, and television transmission
- FF4 Healthcare, medical equipment and drugs
- FF5 Other

To account for outliers, I winsorized the data on a 5% confidence level after the dataset has been imported into R for further analysis. The effect of winsorization on the data distribution is depicted in Appendix 5 Given the data availability and the assumptions made I dropped Switzerland as an observation country since after data wrangling no treatment group transactions remained. For the same reasons, no transactions in 2013 and 2014 remained. After this data manipulation for cleaning purposes was completed, the datasets contained 406 treatment group deals and 2.062 control group transactions carried out between 2013 and 2019. An overview of the final dataset is given in Table 1.

With this information as a starting point of the data sets' structure, the next section will start with descriptive analyses, already partially addressing the research hypotheses, before section 4.4. will use more in-depth statistical procedures to postulate causal relationships and answer the research question.

[77] See, for instance, Graf et al. (2009, p. 2).
[78] See Kaplen et al. (1989), p. 235; Guo et al. (2009, p. 51).
[79] See Wickham (2014, pp. 2-5).
[80] See Harris et al. (2013, p. 7)
[81] See European Commission.
[82] See Hahn (2009, p. 24).
[83] See Beuselinck, Elfers, Gassen, and Pierk (2021, p. 10).
[84] See French's website for more detailed information on industry classification.

Table 1: Treatment and Control Group Characteristics

	Treatment Group	Control Group
FF1	125	503
FF2	88	459
FF3	43	298
FF4	13	78
FF5	136	716
Austria	3	19
France	94	465
Germany	37	242
Italy	123	666
UK	149	670
2015	76	458
2016	102	439
2017	92	496
2018	123	599
2019	13	70

Characteristics of treatment and control group transactions: main industry based on *Fama French* five industries classification, target country and buyout year for the cleaned treatment and control group dataset.

4.3. Descriptive Statistics

My two data sets will be introduced with an overview of KPI levels before comparing treatment and control group transactions in this section. Finally, I will also display first findings on KPI development throughout the observation period. On average, a PE target company has a pre-buyout (T-1) EBITDA of 10.6 Mio. EUR and FCF of 7.4 Mio. EUR while the control group seems to have a lower EBITDA (mean 6.7 Mio. EUR) and FCF (mean 5.2 Mio. EUR).

In addition, a more detailed overview of KPIs for both treatment and control group throughout the observation period is graphically displayed in Appendix 2 as well as presented in Appendix 3. As one can see from this overview, the KPIs driving value creation through operational engineering as defined above do appear to differ. Not only in terms of differences between treatment and control group, but also in terms of skewness - the clear discongruity between median and mean as seen above is only a first indicator for diverging selection patterns between treatment and control group. Besides this, PE transactions also seem to be larger in size (measured by lnAssets) than the control group (mean of lnAssets in T-1 was at 10.41 for the treatment group and 8.09 for the control group). These findings are in line with other author's findings and could give an initial indication to confirm hypotheses one and two.[85]

It has become apparent that the key parameters considered do differ pre-buyout. From this, however, I cannot deduce a significant indication for PE target selection patterns. Therefore, I first apply a t-test on means between treatment and control group transactions.[86] However, the data distribution violates the tests prerequisites of homoscedasticity and normality.[87] In fact, the dataset retrieved from Orbis appears to be comparable to the one used by Acharya *et al.* in distribution as the KPIs are not normally distributed and are left-skewed as well as leptokurtic.[88]

Given the data sets' peculiarities, I conduct Wilcoxon's signed-rank test to test for differences in the median between test and control group, as this test is less sensitive for outliers than a *regular* t-test on means.[89] Setting the treatment group KPI means as the test variable, the null hypothesis of the medians being sufficiently similar can be rejected on a five percent confidence level for all KPIs at least once in T-1 or T0. To assess the strength of the effect, I calculate *Cohen's D* as a measure for effect size.[90] Overall, treatment group transactions appear to be significantly larger in terms of EBITDA, FCF, and lnAssets throughout the observation period, as can be seen in Table 2.

It becomes apparent that the KPI characteristics do differ significantly for most observations. In fact, PE target companies seem to be larger in terms of EBITDA, FCF and assets. Also, unlike control group transactions, leverage appears to increase for treatment group transactions, which is plausible due to the LBO structure.[91] To be able to carry out the analyses to follow in chapter 4.5., I first try to capture tendencies in KPI development for both control and treatment group sep-

[85] See Acharya et al. (2009, p. 15-17).
[86] See Student (1908, p. 1).
[87] The prerequisites were tested using Levene's Test for equal variances to test for homoscedasticity and Shapiro Wilk Variance Test for normality, see Levene; Shapiro and Wilk (1965)
[88] See Acharya et al. (2009, p. 16).
[89] See Dalgraad in Introductory Statistics with R (2008), p. 99. and Acharya et al. (2009, p. 16).
[90] See Cohen (1988, p. 20-21).
[91] I.e., more debt is taken on in T+2 for financing of additional net present value positive projects to, for instance, implement market expansion strategies developed together with the PE firm.

Table 2: Wilcoxon Signed Rank Test on Median

KPI Median	Wilcoxon Signed Rank Test on Median			
	T-1	T0	T+1	T+2
EBITDA	4,428.27*	4,291.80*	4,660.52*	4,051.80*
	*1,614.38**	*1,417.40**	*1,594.69**	*1,468.07**
	(0.29)	(0.27)	(0.25)	(0.27)
FCF	3,129.98*	2,477.78*	998.54	1,552.34
	*1,304.26**	*848.09**	*974.63*	*875.45*
	(0.15)	(0.17)		
lnAssets	10.19*	10.34*	10.57*	10.60*
	*7.91**	*7.96**	*8.06**	*2.65**
	(0.95)	(0.99)	(1.00)	(1.01)
Leverage	42.42%	40.86%*	38.80%	57.46%*
	41.21%	*33.97%**	*33.60%*	*33.14%**
		(0.13)		(0.08)

Wilcoxon Signed Rank Test on the median of treatment and control group. The numbers represent the median for EBITDA and FCF in TEUR. Significant differences in median on a 5% confidence level are denoted with an asterix. Figures in *Italics* state the values for the control group. Added in parentheses is the effect size, if the difference is significant, measured by *Cohen's D*.

arately. I therefore compare the development of KPI growth between treatment and control group transactions throughout the observation period. The results are presented in Table 3.

From this it becomes clear that treatment and control group transactions do not only differ significantly in their pre-buyout characteristics, but they also evolve differently throughout the observation period.

From the initial analyses performed in this chapter, it became clear that in fact there is a significant difference value driving KPIs both pre-buyout as well as afterwards. Besides this, I could also find initial evidence on pre-buyout differences between PE-backed and control group transactions. All the above does in fact suggests initial evidence for the research hypotheses postulated above. For this reason, the next chapter will address these questions in more depth, using a virtually pioneering approach in the sphere of PE research: matching treatment and control group transactions via propensity score matching (PSM) based on their pre-buyout KPI characteristics.

4.4. Research Design

To fully address the hypotheses and ultimately ascertain causal relationships, I will illustrate the methods applied, mainly propensity score matching, in this chapter before section 4.5. will present the results. Overall, PSM describes the matching of two populations using propensity scores (PS) estimated by a logistic regression model. While this approach is a standard procedure in scientific areas where observational studies are predominant (i.e., psychology or medicine), Acharya et al. (2009) were, to the best of my knowledge, the first authors applying this method to performance driver quantification in PE investments.[92] This strategy is particularly intriguing as it introduces new approaches to an existing academic discourse: while there is numerous contributions on PE target selection patterns and PE target performance post-buyout, this method incorporates both streams of literature.

Using this approach, I can compare KPI development with very similar pre-buyout characteristics and a comparable PE buyout likelihood, expressed by the estimated PS.[93] Consequently, I can investigate the effects of PE ownership in comparison to the control group transactions. In addition, PSM incorporates further benign characteristics: as I can reduce selection bias, amongst other biases associated to covariates, by applying PSM in combination with an effective matching algorithm, I can testify relationships without having to consider potential shortcomings weakening my analyses' testimonies as extensively.[94] Furthermore, matching based on the calculated PS allows me to assume the groups to be sufficiently alike and matched transactions to be interchangeable between treatment and control group. Precisely this exchangeability is crucial for causal inference and thus for me to derive causal and statistically significant conclusions from the analyses to be performed in the next chapter. This interchangeability therefore also allows me to presume adequately similar KPI characteristics between control and treatment group. Although the number of control group transactions is noticeably larger, this substitutability in combination with the above unveiled statistically significant inter-group

[92] See Rosenbaum and Rubin (1985, p. 38); Acharya et al. (2009, pp. 14-22).

[93] The PS can be interpreted as the likelihood of the target company being treated, id est, undergoing a PE-backed LBO.

[94] See Acharya et al. (2009, p. 16).

Table 3: Overview of pre-Buyout KPI Characteristics and Development

	T-1 to T0	T-1 to T+1	T-1 to T+2
EBITDA	-402.62	723.44	322.17
EBITDA Margin	3.46%*	3.40%*	1.60%*
FCF	-7,548.19*	-6,017.26*	-5,767,84*
FCF Margin	-23.98%	-24.13%	-12.80%*
lnAssets	0.16*	0.29*	0.39*
Leverage	-9.20%	-15.79%	-8.28%
Sales	3,460.69*	10,080.45*	14,422.83

KPI development throughout the observation period. Displayed in all cases is the mean growth for treatment group transactions. In addition, t test on differences in means between treatment and control group, denoted with an asterix (*) if significantly different on a 5% confidence level. EBITDA, FCF, and sales in TEUR.

differences for pre-buyout KPIs enables me to draw conclusions from analyses based on the matched dataset.[95]

As the prerequisites of PSM appear to be favorable and fulfilled by my dataset, I determine the difference in means of the pre-treatment covariates as a first step. As already in chapter 4.3., this t-test shows, as expected, a significant difference in covariate means. Thus, I continue by running a logistic regression model on the data with the treatment dummy as the dependent and EBITDA margin, FCF margin, lnAssets, leverage, in the buyout year as the explanatory variables to estimate the PS. In addition, I include an industry classification factor dummy variable. The logit model used is displayed in Appendix 10.

While the FCF margin can be interpreted as the quality of a firm's profits, the EBITDA margin accounts for how efficiently the management utilizes the company's resources to generate a return. Thus, the margins represent how many units of FCF, or EBITDA are generated per additional unit of sales. These return on sales figures are well suited to assess operating performance as, unlike for instance return on assets as another widely used KPI, they are not subject to write-ups and write-downs of assets or changes in reporting mechanisms at the time of buyout; this is also why studies applying a similar approach like this paper rely on these KPIs.[96] By choosing these explanatory variables I can account for several factors simultaneously: Leverage represents potential influences of financial engineering while lnAssets controls for firm size, since smaller companies generally generate higher returns, thus being associated with a higher risk of default. Lastly, the introduced margins act as a link between size and returns given they are scaled on sales and should therefore be comparable within peers.

Having performed the underlying logit model, the region of common support of propensity scores for treatment and control group spans from a 0.10 to a 0.94 PS with a mean for the treatment group of 0.68 (control group 0.47) and a median of 0.71 (control group 0.50). The area of common support and PS distribution are also displayed in Appendix 7. The visual inspection once again indicates significant differences in buyout likelihood between both groups: while the control group's PS distribution is evenly distributed with a tendency towards a normally distributed population, the treatment group PS distribution is clearly left skewed. However, a different treatment group PS distribution would, in fact be surprising, given the factual PE involvement.

Based on these propensity scores, a k-nearest-neighbor matching algorithm is executed and assigns sufficiently similar transactions to each other while reducing overall sample-wide distance between PSs. With greedy matching, I receive 400 matched pairs - the remaining control group items would increase overall distance between sample pairs and are therefore discarded off. To assess the quality of the matching algorithm executed, I gauge the PS distribution and the balance of regression covariates. In fact, both PS distribution and covariate balance could have been improved through PSM, as displayed in Appendix 9. From this it becomes apparent that PSM and discarding off unused control transactions did in fact increase similarity within both datasets and the degree of numerical imbalance between the covariates could have been significantly reduced. Thus, PSM was carried out successfully. The indicative results achieved via the regression analyses performed in chapter 4.3 above can therefore be reconfirmed. This can also be seen by the impact of considered KPIs on PS displayed in Appendices 11 and 12.

Based on the matched dataset created through PSM, I will set up additional logistic regression models to evaluate:

i. Differences in pre-buyout characteristics to define a set of KPIs targeted by PE firms and address hypothesis one
ii. KPI development after buyout dependent on group affiliation regarding hypothesis two
iii. Significant influence of operational engineering on KPI and profitability growth in PE-backed transactions to answer hypothesis three

Following the approach of COHN *et al.* (2021), I will set up multiple models controlling for specific characteristics

[95] See Rosenbaum and Rubin (1985, p. 33).
[96] See Cohn et al. (2022, pp. 265-266).

that might influence FSLIs.[97] Supplementing the approach to determine PSs, I will take fixed effects into account by controlling for industry, target home country, and firm age at buyout.

COHN *et al.* also state that there are two main reasons for post-buyout performance in LBOs: either due to unlocking growth opportunities by injecting capital or by distressing struggling firms.[98] For this reason, I will calculate 25% performance quartiles (Q) and investigate whether dependencies of operational engineering do in fact drive value creation in companies based in European. I will therefore investigate time persistence of inter-group quartiles between treatment and control group throughout the observation period.

After having carried out logit models and PSM to determine PE target selection patterns, I will use these results to evaluate the post buyout KPI development dependent on PE ownership through another regression setup. To finally determine whether PE ownership significantly influences value creation through operational engineering, I will use a difference in difference (DiD) regression approach where I will use EBITDA and FCF margin as the dependent variables and add additional explanatory variables. Given this setting, I can control for two-way fixed effects and thereby exclude effects on EBITDA and FCF originating from other sources like financial or governance engineering effects. In addition, I will control for country-, year-, industry-, as well as firm-fixed effects. Thereby I can identify the true effect solely attributable to effects arising from operational engineering.

To address the hypotheses derived in chapter 4.1, the results of my analyses will be presented and expounded in the next chapter. To corroborate my results, chapter 4.6 will critically review the findings and perform robustness tests and sensitivity analyses to critically review the analyses performed.

4.5. Results

Before assessing value creation mechanisms, I consider and analyze pre-buyout characteristics and post-buyout development in sections 4.5.1. and 4.5.2., before applying a DiD approach in section 4.5.3. to account for two-way fixed effects to determine the impact of PE ownership solely attributable to operational engineering.

4.5.1. Analysis of Pre-Buyout Characteristics

To address hypothesis one, I will investigate pre-buyout characteristics of PE transactions to develop a framework of a favorable KPI set for PE transactions. To do this, I perform several logistic regression models do determine effects of KPI levels on buyout likelihood. Overall, I construct six regression models. As shown by other authors, the relevance of certain KPIs might differ depending on their relative size when compared to peers.[99] Therefore, I have also included KPI quartile indicators as explanatory variables. Model I only considers EBITDA and FCF quartile assignment. In model II, I assess pre-buyout EBITDA metrics. To expand this approach, I add lnAsset and Sales quartiles as size proxies as well as leverage as explanatory variables in model III. Models IV and V follow an equivalent setup as regressions II and III, using FCF pre-buyout characteristics instead of EBITDA as explanatory variables of interest. Finally, model VI unites the previous ones considering both EBITDA and FCF margin simultaneously. Including both EBITDA and FCF margin quartile variables in one model is not possible due to the data structure resulting in concerns regarding multicollinearity. The logit models' output is displayed in Table 4.

In all cases, the KPIs as explanatory variables are regressed against the treatment dummy variable, equaling one for PE transactions and zero for control group elements. The model output thereby can be interpreted as the change in likelihood of PE engagement given a change in pre-buyout KPIs. The regression equations are presented in models (A2) and (A3), as shown in Appendix 10.

From these analyses performed, it becomes clear that pre-buyout FSLI characteristics do have a significant impact on PE buyout likelihood - as expected. In particular, EBITDA margin and quartiles as well as sales quartiles as a size measure drive these effects: while higher EBITDA quartile assignment increases the buyout likelihood, above-average profitability appears to have the opposite effect. On the other hand, results on FCF impact are more blurred, as can be seen in models IV and V.

Besides this, PE firms seem to target comparably small firms, measured by sales, as the sales quartile coefficient is significantly negative in all cases. The direction and significance of the effects observed does not change when adding additional explanatory variables worthwhile considering like lnAssets and leverage.

To re-evaluate the results, I have additionally controlled for country-, year-, and industry-fixed effects in separate models, yielding the same results as displayed. Also, the coefficient of determination, expressed by Nagelkerke's pseudo R^2,[100] shows sufficiently high levels of explanatory power for most models. Besides the coefficient of determination, I calculate the root mean squared error (RSME) for each model. With the results achieved, I can confirm the findings of coefficient significance and satisfactory explanatory power for the logit models. However, I could not include margin quartile explanatory variables in model VI due to the dataset's structure and coefficient correlation. Moreover, to mitigate possible concerns regarding explanatory power and model reliance due to correlation within the explanatory variables, I calculate a variance inflation factor (VIF) for all explanatory variables in models I to VI.[101] From this analysis, I can preclude potential model deficiencies arising from multicollinearity.

[97] See Cohn et al. (2022, p. 276).
[98] See Cohn et al. (2022, p. 271); Acharya et al. (2009, p. 2).
[99] See Cohn et al. (2022, p. 262, pp. 260-270).
[100] See Nagelkerke (1991, p. 1).
[101] See Johnston, Jones, and Manley (2018, pp. 1958-1959).

Table 4: Logistic Regression Models on Buyout Probability given KPI Levels at Buyout

KPI	**Logistic Regression Models**					
	I	II	III	IV	V	VI
$EBITDA_{T-1}$ Margin		-0.01**	-0.06^{x}			-0.07**
$EBITDA_{T-1}$ Margin Q		-0.06***	-0.02			
$EBITDA_{T-1}$ Q	0.07***	0.07***	0.09**			0.07*
FCF_{T-1} Margin				0.12	-0.06	0.01
FCF_{T-1} Margin Q				0.07*	0.02	
FCF_{T-1} Q	0.00			-0.09**	0.01	0.02
$lnAssets_{T-1}$ Q			-0.01		0.07	-0.02
$Sales_{T-1}$ Q			-0.20***		-0.18***	-0.19***
$Leverage_{T-1}$ Q			0.03		0.04^{x}	0.03
Pseudo R^2	7.26%	16.73%	19.10%	2.99%	15.75%	19.34%
RSME	0.36	0.45	0.49	0.46	0.46	0.45

Logistic regression outputs predicting buyout probability given level of EBITDA and/or FCF indicators at buyout. Logit I addresses the effect of overall EBITDA and FCF size (quartiles), while Logit II covers different EBITDA characteristics, only. Logit III adds lnAssets and sales quartiles as size proxies and leverage to account for financial engineering. Models IV and V are analogue to models II and III, investigating FCF instead of EBITDA. Model VI investigates both EBITDA and FCF while also controlling for further influencing factors. In each regression, the dichotomous Treatment Dummy variable, taking one for PE buyouts and zero for non-PE backed transactions is the dependent variable. The level of significance is represented by an asterix where the explanatory variables are statistically significant at a 0.1% (***), 1% (**), 5% (*), or 10% (^{x}) confidence level.

To sum up, the findings from the logistic regression models do support the indicative findings as well as the descriptive tests carried out in the previous chapters in terms of quartile effect and profitability and size. Therefore, the next section will focus on KPI development post-buyout to set a starting point on value creation through PE ownership.

4.5.2. KPI Development throughout the Observation Period

As in particular quartile explanatory variables showed very high levels of significance, I want to further evaluate the relevance of KPI quartile assignment and quartile differences between treatment and control group.

To do this, I first perform an analysis of variance (ANOVA) on quartile KPI levels, which yields highly significant differences in means between control and performance group quartiles. Thus, I once again apply a t test on means on each quartile bracket of control and treatment group in the pre-buyout year as well as the end of the observation period in T+2, as displayed in Appendix 6. In line with the results obtained in the logit models, the quartile and margin means do differ significantly between treatment and control group both pre- and post-buyout. In most cases, the above-median companies showed higher levels for all KPIs in the control group. This in in line with the results obtained in the previous section: albeit PE firms appear to target companies with relatively high levels of EBITDA, higher relative levels of pre-buyout sales as a size proxy significantly decrease buyout likelihood. As the analyses carried out so far show similar and statistically significant results, I can already address hypothesis one and hypothesis two partially:

> *PE firms target small firms compared to control group transactions. This can be seen by an on average significantly lower sales base. However, these PE targets seem to be less profitable, as can be seen by the regression results for the included profitability quartiles, namely EBITDA margin and margin quartiles.*

As can be seen, the pre-buyout characteristics do differ significantly as PE firms seem to systematically target potential portfolio companies with a predefined set of KPIs. However, I first and foremost want to evaluate whether PE activity also has a positive impact on these KPIs during the holding period.

To further investigate the initial findings on time persistent differences in KPI quartiles, I perform propensity score matching. In this setting, PSM is a very powerful tool as it allows me to analyze similar companies in terms of pre-buyout characteristics and thereby assess the real impact of PE ownership. To do this, I use a comparable model to the ones displayed in Table 4 to calculate the propensity scores for each transaction and match each treatment group observation to one non-PE backed transaction.[102] The initial results of PS distribution in the new dataset generated through PSM support the findings of the logit models already carried out.

Thus, the analyses performed so far show significant impact of pre-buyout KPI levels on the likelihood of PE engagement. Also, I have demonstrated that relative FSLI size

[102] See Acharya et al. (2009, pp. 21-22, p. 42).

in terms of KPI quartile assignment is persistent throughout the observation period. Specifically, relatively low levels of EBITDA (margin), which is one of the favored pre-buyout characteristics for PE backed transactions, provide opportunities for value generation through operational improvements. The findings of persistent KPI quartile assignment differences between treatment and control group firms also indicate that PE firms do create surplus economic value through operational engineering. These implications of operational engineering measures implemented by the PE firm in fact increasing EBITDA and FCF are thus reinforced by the quartile assignment development as shown above. These findings are also in line with other papers. For instance, COHN *et al.* also find evidence for PE firms targeting comparably unprofitable firms, as this allows the highest potential for margin improvement, what the authors refer to as "*turnaround opportunities*".[103]

However, from both, the findings on pre-buyout characteristics as well as the significant differences in quartile assignment throughout the observation period, no causal relationship between operational improvements implemented by the PE firm and KPI enhancement as well as overall higher profitability improvements can yet be drawn. Albeit these findings may suggest a significant relationship, changes in KPIs as well as their underlying FSLIs can just as likely be due to other reasons. For instance, leverage used by the PE firm may lead to more capital readily available in the first periods after buyout that could be used to launch new products or enter new markets and thereby increase sales and thus EBITDA, as interest payments are not incorporated in this figure. While this would be associated to financial engineering, also increased efficiency through improved governance mechanisms may yield higher EBITDA or FCF. However, in the presented analyses, this would spuriously be assigned to operational engineering effects. Therefore, I cannot yet deduce a causal relationship between PE involvement and increasing EBITDA and FCF solely attributable to operational engineering from the results obtained so far.

4.5.3. Difference in Difference Analysis of Private Equity Ownership Effect on Operational Engineering

To establish a causal relationship of whether PE firms make use of operational engineering to increase profitability and thereby generate value, this chapter will use methods capable of determining causal inference.

Besides operational and financial engineering as the two performance driving strategies resulting in a direct bottom-line effect, also time-, industry-, country-, and firm-fixed effects likely pose a relevant factor in KPI development. However, as most of these effects do influence the same KPIs and FSLIs, there are interdependencies between all of them.

To finally address hypothesis three and the overarching topic of this paper, I will decompose the growth effects originating from macroeconomic effects, leverage, and operational improvements. For this reason, I will use a multivariate analysis that can capture two-way fixed effects. The interacting two-way fixed effects, namely the simultaneous influence of pre- and post-buyout characteristics as well treated and untreated item-fixed effects can be analyzed in a difference-in-difference (DiD) setting. By adding lnAssets as a size proxy I can control for firm size effects while leverage as another explanatory variable captures profitability gains through financial engineering. In addition, by including *Fama French* industry factor dummy variables, I can control for industry-wide time-series variation in business conditions.[104] From this, I can genuinely assess the value generation attributable to operational engineering without neglecting effects arising from other sources like financial engineering or firm- and industry specific circumstances. To perform this DiD regression, I first manually transform the cross-sectional dataset retrieved from Orbis and transformed through PSM into a panel data set.

After this final data preparation, I set up two linear regression models, one for EBITDA and FCF margin, contemplating every transaction i in every period t. The models read as follows:

$$\begin{aligned} EBITDA_{Margin(i,t)} &= \beta_0 + \beta_1\left(Treatment_{i,t}\right) \\ &+ \beta_2\left(postbuyout_{i,t}\right) + \beta_3\left(Treatment \times Post\right) \\ &+ \beta_4\left(Leverage_{i,t}\right) + \beta_5\left(lnAssets_{i,t}\right) \\ &+ \beta_6\left(Industry_{i,t}\right) + \varepsilon_{i,t} \end{aligned} \tag{1}$$

$$\begin{aligned} FCF_{Margin(i,t)} &= \beta_0 + \beta_1\left(Treatment_{i,t}\right) \\ &+ \beta_2\left(postbuyout_{i,t}\right) + \beta_3\left(Treatment \times Post\right) \\ &+ \beta_4\left(Leverage_{i,t}\right) + \beta_5\left(lnAssets_{i,t}\right) \\ &+ \beta_6\left(Industry_{i,t}\right) + \varepsilon_{i,t} \end{aligned} \tag{2}$$

In this setting, the influence of several factors on EBITDA and FCF margin can be assessed simultaneously. Besides the regression's intercept β_0, β_1 as the first DiD component displays the overall PE ownership effect. This entails the pre-buyout period T-1, as the exact date of transactions is not taken into consideration due to data availability. In addition, β_2 indicates the development as of T0 for all transactions and represents the second DiD component - post buyout. Finally, β_3 unites both DiD aspects by adding explanatory power on the effect of PE ownership on value creation, which ultimately is the variable of interest.

In addition, potential influences arising from financial engineering are considered by the regression coefficient β_4. In fact, the post-buyout PE-ownership value creation factor β_3 can describe value creation solely attributable to operational engineering measures. Additionally, lnAssets and *FF* industry classification factor dummy variables are added as covariates, acknowledged with coefficients β_5 and β_6. In an additional model I have controlled for unobserved confounders

[103] Cohn et al. (2022, p. 271).

[104] See Cohn et al. (2022, p. 271).

Table 5: Difference-in-Difference Regression Models - Influence of PE ownership on Profitability through Operational Engineering

KPI	**DiD lm Regression Models**					
	I	II	III	IV	V	VI
β_1 Treatment	-0.29***	-0.29***	-0.29***	-0.03	-0.09	-0.09
β_2 Post Buyout	-0.28***	-0.26***	-0.27***	-0.26**	-0.25**	-0.25**
β_3 Post Buyout Treatment	0.29***	0.27***	0.28***	0.03	0.05	0.05
β_4 Leverage	0.00***	0.00**	0.00**	0.00	0.00	0.00
β_5 lnAssets		-0.02*	-0.02*		-0.07***	-0.06***
R^2	5.15%	5.41%	5.50%	3.06%	5.46%	5.62%
RSME	0.56	0.56	0.56	0.58	0.57	0.57

DiD Regression with PSM EBITDA and FCF Margin as the dependent variables. DiD model with two-way fixed effects also taking treatment point of time T0 into consideration with parallel observation of influence of explanatory variables on dependent variable as measure for operational engineering quality. Models I to III display output with EBITDA margin as dependent variable, models IV to VI with FCF margin. While models I and IV only include the treatment dummy and leverage to assess the impact of PE ownership and parallel impact of financial engineering through leverage, models II and V also control for firm size using lnAssets as a proxy. Models III and VI also control for *Fama French* industry-fixed effects. The level of significance is represented by an asterix where the explanatory variables are statistically significant at a 0.1% (***), 1% (**), 5% (*), or 10% (^{x}) confidence level.

by including country-, year-, and firm-fixed effects dummy variables, achieving the same overall results.

I start the analysis by only including the two-way fixed effects coefficients β_0 to β_4 in model I for EBITDA margin. The same model with FCF margin as the dependent variable is displayed in model IV. In models II and V, I also include lnAssets as an explanatory variable. To assess the model quality, I add the coefficient of determination, measured with Nagelkerke's pseudo R^2. In addition, I calculate RSME as a second quality measure. To mitigate potential concerns arising from multicollinearity, I calculate VIFs for every coefficient also in this model.[105] The results from the model quality tests are satisfactory. The results of the DiD regression are displayed in Table 5.

For models I to III describing EBITDA margins, one can clearly see a strongly significant decline post buyout for all models. However, the relevant two-way fixed effects coefficient β_3 is strongly significant and positive in all cases. This coefficient will only be positive for PE-backed firms after the buyout has occurred. Interestingly, this effect becomes astonishingly strong when comparing it to the overall post-buyout development, depicted by β_2: the overall post-buyout development of EBITDA margin turns out to have a negative slope. In comparison to the two-ways fixed coefficient, the effect of PE ownership (treatment group) on this KPI's development turns out to be even stronger. I can therefore conclude from this that PE ownership has s significantly positive influence on EBITDA margin improvement post-buyout. Furthermore, this margin improvement is achieved through operational engineering measures. As I control for effects from financial engineering, namely leverage, and size, with lnAssets as a proxy, as covariates as well as year- and industry-fixed effects in a DiD-setting, this effect can thus solely be attributed to operational engineering measures. In fact, leverage does not seem to significantly impact EBITDA margin, just as industry classification.

In addition, adding more explanatory variables in models II and III (and V and VI respectively), does not increase the explanatory power significantly, as can be seen by a stable coefficient of determination. However, the coefficient of determination shows overall rather low levels. For this reason, I add the root mean squared error (RMSE) for all models to evaluate their overall fit. Like the coefficient of determination, the RMSE does not change significantly when adding additional explanatory variables. Therefore, the combination of highly significant regression coefficients with sufficiently low RMSEs represent strong analytical evidence.

In contrast to the findings on EBITDA margin, the models assessing PE impact through operational engineering on FCF margins show comparable results, thus not being as reliable in terms of statistical significance. Only the post buyout coefficient, just as in the EBITDA margin models, turned out to be significantly negative in model IV. While the direction of the post-buyout treatment and two-ways fixed effects coefficient β_3 is the same in models IV to VI, they are smaller in absolute size - and insignificant. However, the initial FCF DiD model IV also does not convey substantial overall explanatory power, as it yields the lowest coefficient of determination of all six models. Also, the leverage effect did not add significant explanatory power in models IV to VI, while firm size showed a comparable impact on FCF margin as in models I to III on EBITDA margin. Unlike in the EBITDA margin models, however, the introduction of additional covariates β_5 as a firm size proxy and β_6 to account for industry-fixed effects

[105] See Johnston et al. (2018, pp. 1958 - 1959).

does add additional explanatory power to the FCF margin models, achieving comparable explanatory power like models I to III. Thus, even though models IV to VI indicating similar trends compared to models I to III, they did not turn out to add explanatory power to post-buyout treatment effects by PE firms through operational engineering for FCF margin improvement.

With the results obtained from these DiD two-way fixed effects models, I can also address hypothesis two and three, after already having answered hypothesis two above in parts. As demonstrated, pre-buyout characteristics between treatment and control group firms do differ significantly. Furthermore, these KPIs do develop not only at a different pace, but also differently when comparing treatment and control group transactions. This could have been seen through persistence in significantly different KPI quartile allocation throughout the observation period. While this addresses hypothesis two, hypothesis three can be answered with the last analysis' findings. In fact, PE-backed firms do show a significantly stronger increase in EBITDA margin as a profitability measure. As this can be found in a DiD-setting, I can assign this improvement to effects that can be traced back to operational engineering improvements implements by PE firms in their respective portfolio companies. Since operational engineering generally is referred to as EBITDA and FCF effects,[106] I have also investigated FCF margin improvements through operational engineering. While the analyses performed are in line with the findings for EBITDA margin improvement, the lack of coefficient significance does not allow me to assume a causal relationship between operational engineering and FCF margin, unlike with EBITDA margin.

As I could successfully address all three research hypotheses developed in section 4.1., the next section will address potential weaknesses of the analyses performed by carrying out tests on robustness and sensitivity analyses. After the data analysis result reconfirmation sections have been concluded, I will put my findings in an academic framework, comparing my results to comparable papers in section 5.1. Chapter 5.2. will also address potential weaknesses of the analyses performed and will critically review the assumptions made before section 5.3. finally summarizes the findings and chapter 5.4. concludes this paper by demonstrating potential avenues for further research.

4.5.4. Model Evaluation: Tests on Robustness and Sensitivity

In the last section, I have demonstrated and elaborated on the impact of PE involvement on statistically significant improvements in profitability. By applying DiD-models and thereby accounting for two-way fixed effects, I can distinctively assign this margin improvement effect to operational engineering measures. However, while the results from the DiD regressions performed do show significant evidence for EBITDA margin improvement post-buyout, the results are not fully unambiguous given the results regarding FCF margin improvement as well as the coefficients of determination. In view of the lack of verifiably positive impact of PE activity on FCF margin improvement, a commensurable figure to EBITDA margin, the results should be examined more critically. For this reason, I revalidate the model outputs by testing for robustness.[107]

I therefore start by visually inspecting robustness of the models defined in equations (1) and (2) as well as the output. Just as the difference in coefficient significance, the results do differ when comparing EBITDA and FCF margin regression residuals, as displayed in Appendix 13 for models I to III and Appendix 14 for models IV to VI: EBITDA margin shows a sufficiently homogeneous distribution of fitted and residual regression coefficients as well as fit to theoretical vs. actual regression quartiles to test for heteroskedasticity and normality. Similarly, I investigate overall data distribution: EBITDA margin residuals show a right skewed and strongly leptokurtic distribution. In contrast, however, FCF margin regression residuals display a left-skewed residual distribution while also showing some evidence for homoskedasticity and non-normally distributed residuals, which, however, is in line with the findings retrieved through DiD regressions IV to VI.

As the results of visually inspecting the propensity score matched data regression outputs and performing analyses on robustness, I finally want to reaffirm the results by conducting a sensitivity analysis through model variation tests. To do this, I apply a commonly used approach to reducing the present PSM sample to sub-groups.[108] I do this in two steps: first apply the DiD regression model on the dataset while excluding one country per model. As another superordinate model, I control for sensitivity by buyout year. In addition, I control for firm-fixed effects. The regression results for country and industry level sensitivity analyses are displayed in Appendices 15 and 16. As well as the visual inspection as tests on robustness, the sensitivity analyses do confirm the overall significant impact of PE ownership on EBITDA margin improvement post-buyout and thereby operational engineering as a highly relevant value creation driver in PE transactions.

5. Discussion

Having presented the analyses results in the previous sections, I will now summarize and discuss my findings. I start by putting my results in a framework of the current academic discourse. Subsequently, I will discuss my results in chapter 5.1. and compare them to other authors findings on PE value creation through operational engineering. From this, I will draw a conclusion and assess the implications of my results. Thereafter, I will critically review my results and section 5.2. and discuss potential weaknesses of the models applied and analyses presented. I will sum up this paper with concluding remarks in chapter 5.3., before finally showing possible

[106] See, for instance, Achleitner et al. (2010, p. 19); Achleitner et al. (2011, pp. 2-3).

[107] See Lu and White (2014, p. 1).

[108] See Salciccioli, Crutain, Komorowski, and Marshall (1973, pp. 265-267).

avenues for further research derived from my work and the academic discourse.

5.1. Implications of the Results Achieved

In this paper, I have analyzed the effect of private equity ownership on value creation through operational engineering. I provide evidence on pre-buyout characteristics in terms of PE firm target selection as well as KPI improvements after buyout through operational engineering. I have applied logistic regression models to determine buyout likelihood given a set of relevant KPIs. From this analysis I can conclude the following buyout characteristics favored by PE firms as

reflected in my sample: in comparison to the control group transactions, PE firms target small firms measured by sales. The targeted companies are comparably unprofitable regarding EBITDA margin. I have reaffirmed these characteristics by determining the KPI quartiles and their development throughout the observation period, finding statistically significant and time persistent differences between treatment and control group.

Following the pre-buyout characteristics, I have demonstrated the relevance of operational engineering activities for value creation in PE transactions. I have shown this by applying propensity score matching and thereby comparing extraordinarily similar companies. By using a DiD approach, I have controlled for two-way fixed effects of year- and industry-fixed effects as well as interdependencies resulting from financial engineering on the KPIs to be evaluated. From this analysis, I can conclude that PE firms are particularly effective in applying operational engineering activities to increase profitability. The results achieved keep their overall explanatory power when testing for model robustness by adding additional explanatory variables and have been reaffirmed by performing sensitivity analyses.

As extensively highlighted in the introductory sections, operational engineering represents, amongst financial and governance engineering, one major driver in value creation in PE transactions. Following the mathematical decomposition introduced by Achleitner *et al.* with the value creation bridge, EBITDA and FCF effect can be described as the main drivers yielding surplus value created through operational engineering.[109] Since value creation through operational engineering can also be perceived as the metrics attained through actual measures and skillful implementation of successful actions by the PE firm, I have focused on these KPIs within my European deal-level data set.

In line with COHN *et al.*, who state to be the first to determine PE target characteristics and predict favorable KPI sets of companies PE firms acquire, I have performed comparable analyses on my dataset.[110] I also find a significantly negative impact of relative size, measured by sales quartiles, on buyout likelihood. While I find the tendency of PE firms to target comparably unprofitable firms, the authors postulate a U-shaped relationship with higher buyout likelihood for both extrema of (un-)profitability. While I cannot fully reconcile these findings with my dataset, I can partially support this statement as EBITDA margin has a significantly negative impact on buyout likelihood, while higher EBITDA quartile assignment has significantly positive impact on buyout probability.

COHN *et al.* also elaborated on two distinctive theories on why PE firms may be attracted by highly (un-)profitable companies: they either target highly profitable firms because of *"untapped growth opportunities because of financial constraints"*[111] or unprofitable firms as these companies could serve as a growth platform with extraordinarily large optimization opportunities to capitalize on.[112]

This also found by Achleitner *et al*., stating that high profitability pre-buyout is not associated with larger margin improvements during the holding period.[113] With the data collected from Orbis, I can only find evidence on comparably unprofitable target companies in terms of EBITDA margin, supporting the hypothesis of PE firms aiming at the acquisition of companies where they fully use their knowledge and capabilities to increase margins in low-performing firms to capitalize on, which is also what Stafford finds for his dataset. In addition, he also finds evidence on PE firms targeting small firms.[114] Besides size, the academic findings on relevance of leverage on buyout likelihood are inconclusive. While some authors postulate evidence on the relevance of leverage and its decrease during the holding period,[115] Stafford and other authors, just as I, find no evidence on leverage being a highly relevant KPI predicting buyout likelihood.[116] However, as Stafford uses public-to-private transactions, the mean firm size in the dataset likely is larger and thus PE transactions might appear to be relatively small in comparison to the other transactions included in his dataset. Nevertheless, this might also be the case for my dataset - this could be assumed given the significantly higher mean sales volume for control group transactions.

On the other hand, however, Acharya *et al.* cannot confirm these findings as they find evidence of the selection pattern being non-linear in profitability, thus PE firms targeting companies that are neither unprofitable nor highly profitable.[117] My results achieved through PSM partially support this view, as well, as displayed in the margin PS distribution displayed in Appendix 11.

Comparing the results presented by Cohn *et al.* and Acharya *et al.*, I would classify my results as a finding at the intercept of both papers: while I cannot find evidence for PE firms targeting firms with above-average EBITDA margins, as found by Cohn *et al.*, this does not necessarily imply that the

[109] See Achleitner et al. (2010, p. 19).
[110] See Cohn et al. (2022, p. 260).
[111] Cohn et al. (2022, p. 271).
[112] See Cohn et al. (2022, pp. 268-270).
[113] See Achleitner et al. (2011, p. 14).
[114] See Stafford (2015, p. 12).
[115] See Achleitner et al. (2010, p. 5).
[116] See Stafford (2015, p. 11).
[117] See Acharya et al. (2009, p. 5).

margins are negative, as can also be seen by the KPI quartile means calculated. This supports the view of Acharya *et al.*, who also postulate that PE firms target companies with high upside, but low downside potential. This pattern could also explain the results achieved in my logit models and can, as discussed, be explained by the high due diligence efforts usually entailed in PE transactions.[118]

As the analysis on pre-buyout KPI characteristics and the derivation of a KPI level set favored by PE investors has been carried out successfully, I have further followed the approach of Cohn *et al.* and Acharya *et al.* by performing PSM to create dataset of treatment and control group transactions which are highly similar in pre-buyout KPI characteristics. Analogue to these contributions, I perform more advanced statistical analyses on the propensity score matched dataset to assess the impact of PE involvement on post-buyout KPI development. In fact, my finding of increased profitability in PE-backed transactions after buyout, as unveiled through DiD two-way fixed effects regression, has also been found by Cohn *et al.*: their result of PE involvement significantly increasing profitability for propensity score matched peers, which is even stronger the lower the profitability pre-buyout, can also replicated with my dataset and analysis.[119] I was able to demonstrate that companies with low levels of EBITDA margin pre-buyout for PE targets grow significantly stronger by operational engineering measures. Thereby I can convey the same testimony as Cohn *et al.* have by stating that PE firms also target firms with lower profitability as they are capable *"to turn around struggling firms"*.[120]

Thus, as a preliminary conclusion, I can summarize that PE firms use a defined set of pre-buyout KPIs for potential portfolio companies and, unlike non-PE backed firms with extremely similar FSLI characteristics, significantly increase profitability. They do this by operational engineering, primarily addressing EBITDA and thus the main drivers for operational engineering.[121]

The results of Cohn *et al.* as well as mine are also in line with the results of Acharya *et al.* on operational improvements. This paper also shows evidence on gains in profitability through operational improvements for PE-backed transactions. In fact, my results of overall margin improvement versus post-buyout PE impact on EBITDA margin as demonstrated in the DiD regressions can also be compared to the analysis performed by Acharya *et al.*: while I could not find evidence on EBITDA margin gains for the overall PSM dataset, PE engagement showed a highly positive and significant impact post-buyout. Also, with the analyses and additional tests and taking two-way fixed effects into account, I can assign these gains in EBITDA margin improvements to operational engineering.[122] What's even more, my findings of PSM increasing the positive impact of PE ownership on above-average profitability improvements, in comparison to the overall dataset, is what Cohn *et al.* could also find.[123]

Besides the concurrence of this papers' findings with existing literature applying PSM, Hahn also found significant impact of PE ownership on value creation through operational engineering in a DiD setting comparable to mine. In fact, he also found a significant post-buyout treatment effect on EBITDA margin growth resulting in abnormal performance due to the relevant KPIs *"being causally altered by PE ownership"*.[124]

By taking two-way fixed effects into consideration when assessing the relevance of operational engineering, I have also considered the relevance of leverage, thus financial engineering, on buyout likelihood. The development throughout the observation period and results from the analyses performed suggest lower relevance of financial engineering on value creation. This is also what Cohn *et al.* found in more thorough investigations.[125] Even though I could not confirm findings on FCF effect with my dataset, I could prove the relevance of EBITDA effect as the primary driver of value engineering operational engineering.[126] Overall, the findings presented in this paper are in line with the prevailing sentiment in academic discourse.[127]

5.2. Potential Weaknesses and Shortcomings

So far, I have covered the strengths of my analyses and have put their implications in the context of other author's contributions. Albeit having conducted tests on robustness and sensitivity analyses, I also want to address potential weaknesses of my analyses and areas of interest not covered in this paper before presenting potential avenues for further research starting points in the last chapter.

First and foremost, in line with academic consensus, I have decided to investigate an observation period of four years in this paper, three of which after buyout (including the buyout year). While collecting the data from the Orbis Bureau van Dijk database, I have only included transactions where I was able to retrieve all relevant KPIs for the entire observation period. These detrimental KPIs are, EBITDA, assets, the FSLIs to calculate FCF according to the definitions displayed in Appendix 1, and leverage. While an average holding period of around four years for PE investments in Europe seems plausible,[128] this may have led to low levels of selection bias as I most likely have excluded several transactions where not every KPI was available for every single year

[118] See Puche (2016, p. 41)
[119] Cohn et al. (2022, pp. 272-273).
[120] Cohn et al. (2022, p. 270).
[121] See Achleitner et al. (2010, p. 19).
[122] See Acharya et al. (2009, p. 12, pp. 24-25).
[123] See Cohn et al. (2022, p. 258).
[124] Hahn (2009, pp. 27-28, p. 43).
[125] See Cohn et al. (2022, p. 283).
[126] According to most academic contributions, EBITDA effect appears to be the most relevant value creation driver within the operational engineering strategy. See, for instance, Puche (2016, pp. 40-42).
[127] See Achleitner et al. (2010, pp. 25-26); Achleitner et al. (2011, pp. 14-15); Biesinger et al. (2020, pp. 28-19); Graf et al. (2009, pp. 25-26); Guo et al. (2009, p. 28); Kaplan and Strömberg (2009, pp. 132-133, p. 143); Puche (2016, p. 41).
[128] See Achleitner et al. (2010, p. 25).

- for instance, I have excluded eleven Swiss control group transactions as my preconditions led to all Swiss treatment group transaction being dropped. Also, besides the figures I classified as detrimental for my analyses, data availability was poor for several KPIs. For this reason, I was not able to calculate EV for a sufficiently large subset of treatment and control group transactions. Therefore, I could not translate the findings on the positive impact of PE ownership on profitability into the influence of operational engineering on EV/EBITDA multiple, a widely used multiple in the sphere of private transactions, and thereby quantify the actual value created.

As seen from the first analyses in section 4.2., both treatment and control group transactions do significantly differ in their pre-buyout characteristics. This difference may lead to overt bias, which could occur when, already before treatment, the treated and control group differ in their characteristics. Indeed, I am aware of statistically significant differences between the groups to be compared but nonetheless carry out the analyses. However, by applying PSM, the negative influence of overt bias can sufficiently be reduced as I have applied k nearest neighbor matching as an algorithm reducing overall distance of propensity scores and thereby only taking very similar transactions into consideration for further evaluation.[129] This matching method also is well suited to wipe out potential biases arising from the control group being significantly larger than the treatment group.[130] Therefore, PSM through k nearest neighbor matching is a suitable method to achieve reliable causal inference in my dataset.[131]

After carrying out analyses on probability of PE involvement based on pre-buyout FSLI characteristics, I have carried out DiD regression models to assess the impact of PE ownership on operational engineering. As discussed in the previous section, my findings of PE ownership resulting in significantly higher EBITDA margin improvements is in line with other author's findings. This setup is capable of assessing the implications of operational engineering, while accounting for two-way fixed effects, namely hidden effects of financial engineering and year-as well as industry-fixed effects. In addition, I have also controlled for country-, year-, and firm-fixed effects through the sensitivity analyses. Also, using a linear regression model as standard method that is widely used is favorable as other methods in R entail very specific prerequisites and are not necessarily similarly well suited for my analyses.

While the statistical methods applied to analyze the data at hand were proven to be suitable, the output might in parts provide reasons for doubt. In line with Cohn *et al.* and Acharya *et al.*, amongst others, I find evidence on profitability gains through operational engineering.[132] As postulated and mathematically decomposed by Achleitner *et al.*, EBITDA and FCF are the main operational engineering value creation drivers. While I found strong and consistent evidence on improvements in profitability measured by EBITDA margin, this was not the case for FCF margin. The post-buyout effect of PE ownership on FCF margin was positive, nevertheless not significant in my models. Also, I could not find evidence for FCF margin improvement as a relevant source of value creation through operational engineering. What is interesting is the fact that when examining buyout probabilities, the impact of FCF margin pre-buyout appeared to have statistically insignificant thus opposing effects on probability of PE involvement.

To reaffirm the results obtained, I have performed additional tests to assess their significance. Overall, the tests on robustness confirm the relevance of EBITDA as a highly relevant performance driver in value creation through operational engineering. However, just as in the DiD models, I could not find genuine evidence for FCF margin improvement.

Besides tests on robustness and sensitivity analyses, I have mainly assessed model quality by interpreting the coefficient of determination. Given their values being rather low in the DiD setting, I critically reviewed my analyses. However, when comparing my results to the ones of Acharya *et al.*, my results appear to be comparably good and thus sufficiently strong in explanatory power, as they find coefficients of determination between 6% and 16%.[133] In addition, I have calculated RSME as additional model quality assessment and VIF to mitigate concerns regarding variable correlation.

5.3. Conclusion

In this paper, I have investigated value creation mechanisms through operational engineering, the driver gaining more and more relevance for PE firms to create excess economic value in last decades.[134] In line with recent literature, I have focused on deal-level data of Europe-based target firms, following and uniting approaches presented in recent academic contributions in my analyses.[135] With the analyses performed and put in an academic framework, I conclude my findings by addressing the research hypotheses developed in section 4.1. as follows:

a. PE firms have a distinct selection pattern. They target firms with comparably low levels of sales volume that are unprofitable measured by EBITDA margin. Also, high levels of leverage do not have a significant impact on buyout probability.
b. PE-backed firms do significantly increase profitability (measured by EBITDA margin). The margin improvement is significantly stronger in PE-backed transactions and time persistent throughout the observation period.

[129] See Rosenbaum (2010, pp. 74-75).
[130] See Ferman (2021, p. 1).
[131] See Stuart (2010, p. 9-10).
[132] See Acharya et al. (2009, p. 24).
[133] See Acharya et al. (2009, p. 41).
[134] See, for instance, Achleitner et al. (2010, pp. 17-18); Harris et al. (2013, p. 20).
[135] See Acharya et al. (2009, pp. 14-22); Cohn et al. (2022, pp. 262-264); Hahn (2009, pp. 42-44).

c. With the analyses performed, I can demonstrate a significant increase in EBITDA margin solely attributable to operational engineering as a prevailing driver of value creation in PE transactions. My findings are in line with academia mainly stating that operational engineering is the primary driver of value creation in recent transactions, especially in Europe.

By considering two-way fixed effects and thereby disentangling simultaneous effects on KPIs, I could reaffirm target selection patterns and KPI development throughout the observation period. By applying PSM, I could create a dataset of extremely similar matched transactions which, unlike the overall dataset, did not show any significant differences in FSLIs. Through this, I was able to add additional explanatory power to all models performed. Even when using this dataset where treatment and control group transactions are mutually exclusive as well as collectively exhaustive, I demonstrated significant EBITDA margin improvements post-buyout for PE-backed transactions. By adding a two-ways fixed effects coefficient also controlling for buyout year as well as lnAssets and sales as a size proxy to control for size-fixed effects and leverage to take returns from, i.e., tax shield effects into account, I can decisively define EBITDA effects as a result of operational engineering measures employed by the PE firm, as suggested by Achleitner *et al.*, amongst others.[136] To also consider country-, industry-, year, and firm-fixed effects, I have performed tests on robustness and sensitivity, which reaffirm my overall results.

So far, most existing literature has focused on US and UK based transactions on fund-level data.[137] However, as the European market appears to show different characteristics in value creation, I followed the approach of other authors by applying existent findings and methods to the second largest geographic region for PE transactions: I analyzed European G7 country-based target firm transactions, amended by GSA countries. In addition to the existing approaches on value creation in academia, I have combined two research streams. I use the pioneering approach of applying PSM in a PE setting, like Cohn *et al*. and Acharya *et al*., and apply DiD regressions on this dataset, as suggested by Hahn.[138] To the best of my knowledge, this is the first paper combining these hitherto often overlooked approaches.

My results shed light on margin improvements through operational engineering as a result of PE ownership. They reconfirm existing findings on value creation and combine the benefits of comparing similar PE and non-PE transactions while controlling for two-way fixed effects for transactions in Western Europe. With these results, I can find the same implications of value creation strategies applied by PE firms as other authors. This also means that PE firms do create actual value through their actions. Their impact on margin improvements as examined in this paper is significantly stronger than for non-PE backed transactions. Therefore, I cannot affirm claims such as PE firms not creating value but only transferring wealth through complex compensation schemes (see value transfer hypothesis) and high costs with the only goal of PE firms aiming at realizing swift profits for themselves.[139]

5.4. Avenues for Further Research

While this paper has provided additional evidence on value creation through operational engineering by combining novel approaches, not all relevant factors were in scope and could be covered in the course of this work. For this reason, I want to conclude this paper by presenting interesting opportunities for future work in the sphere of LBOs in general and PE value creation in particular.

First and foremost, data availability did not allow me to calculate EV. For this reason, augmenting my dataset by adding transactions from other commercial databases probably allows doing this and therefore would be worthwhile considering. With an expanded data set, one could translate the impact of operational engineering on overall returns to the GP and even LP to determine, for instance, whether higher EBITDA improvements in PE transactions are reflected in multiple valuation after buyout. Thus, a joint consideration of KPI and multiple development (e.g., EBITDA/EV multiple) would be a promising approach for further research.

While several authors find strong evidence on the relevance of FCF in value creation, my analyses could not find a significant relationship. Therefore, extending this work by further investigation on FCF margin improvement might be useful, too. Also, more in-depth work on pre-buyout characteristics to add more evidence on the selection pattern regarding profitability, given the diverging findings in this special aspect of Cohn *et al.* and Acharya *et al.* and my results, that share characteristics of both analyses, could be interesting.[140]

Given the powerful tool of PSM, more research using this approach is desirable. While academic contributions on PE value creation for European transactions has significantly increased in the last decade, the overall understanding of value creation mechanisms in this market is not yet as mature as it is in the US. Therefore, more thorough investigations of the development of value creation mechanisms as well as their implications on overall returns in comparison to Anglo-American transactions using new approaches is an interesting track to follow. Besides this, also a focus on Eastern European transactions could be interesting, as there are barely any academic contributions on these market dynamics, so far.[141] Similarly, only few authors studied value creation and

[136] See Achleitner et al. (2010, p. 9).

[137] This can be seen by the differing value-driving mechanisms. See, for instance, Achleitner et al. (2011, p. 17, pp. 25-26).

[138] See Acharya et al. (2009, pp. 14-22); Cohn et al. (2022, pp. 262-264); Hahn (2009, pp. 42-44).

[139] See Stafford (2015, pp. 26-30); Anders (1992, pp. 8-12); Lowenstein (1985, p.731).

[140] See Acharya et al. (2009, pp. 22-25); Cohn et al. (2022, p. 269, pp. 274-277).

[141] For one academic contribution on value creation through operational engineering see Rikato (2014, pp. 22-45).

selection patterns in PE transactions for Asia-Pacific.[142] Thus, comparative analysis of these markets with geographies like Europe and North America, where the PE industry appears to be more mature, likely is insightful, as well.

It will be interesting to see how the volumes of LBO transactions will change in the years to come with higher macroeconomic uncertainty and rising interest rates and inflation globally. As leverage was particularly high in the last years due to the low interest rates and credit spread,[143] this might change in the next years. In fact, the availability of inexpensive debt will likely decrease which will have an impact on the entire PE market.[144] Thus, another shift from the booming PE industry into other asset classes might reinforce the cyclicality of PE as an alternative asset class. Also, as prices for PE investments were at an all-time high in recent years, institutional investor may reallocate investments for publicly traded securities or other asset classes.

Overall, it remains to be seen whether the current macroeconomic situation has set an end to the buyout boom in recent years.[145] The last buyout waves shifted the relevance of value creation mechanisms from financial to operational engineering. It nevertheless remains to be seen whether new macroeconomic conditions still offer sufficient opportunities to employ these strategies. Considering global developments, it is conceivable that the well-established mechanisms of PE target selection patterns and value creation strategies will alter - whether it will be governance engineering or indeed completely new value driving factors to be employed remains to be seen altogether. The next years will most likely impressively show whether overall returns and value creation strategies have become more resilient and if PE target firm selection patterns adapt to the new situation.

[142] For a comparative analysis of PE value creation in Europe and Asia see Puche (2016, pp. 22-73).
[143] See Acharya et al. (2007, p. 9).
[144] See Achleitner et al. (2010, p. 17).
[145] See PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft (2020, pp. 18-21).

References

Acharya, V. V., Franks, J., & Servaes, H. (2007). Private Equity: Boom and Bust? *Journal of Applied Corporate Finance*, *19*(4).

Acharya, V. V., Hahn, M., & Kehoe, C. (2009). Private Equity Target Selection: Performance and Risk Measurement based on Propensity Score Matching. *Working Paper*.

Achleitner, A.-K., & Braun, R. (2015). *Handbuch Entrepreneurship. Entrepreneurial Finance. TUM School of Management* (G. Faltin, Ed.). Wiesbaden, Springer Gabler.

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

Achleitner, A.-K., Braun, R., Engel, N., Figge, C., & Tappeiner, F. (2010). Value Creation Drivers in Private Equity Buyouts: Empirical Evidence from Europe. *The Journal of Private Equity*, 17–27.

Anders, G. (1992). *The "Barbarians" in the Boardroom. Harvard Business Review.* https://hbr.org/1992/07/the-barbarians-in-the-boardroom.

Asness, C. S., Krail, R., & Liew, J. M. (2001). *Do Hedge Funds Hedge? AQR Capital Management LLC.* https://ssrn.com/abstract=252810.

Axelson, U., Jenkinson, T., Strömberg, P., & Weisbach, M. (2012). Borrow Cheap, Buy High? The Determinants of Leverage and Pricing in Buyouts. *Forthcoming Journal of Finance*.

Axelson, U., Jenkinson, T., Weisbach, M. S., & Strömberg, P. J. (2008). Leverage and Pricing in Buyouts: An Empirical Analysis. *Working Paper*.

Berg, A., & Gottschlag, O. (2003). Understanding Value Generation in Buyouts. *Journal of Restructuring Finance*, *2*(1), 9–37.

Beuselinck, C., Elfers, F., Gassen, J., & Pierk, J. (2021). Private firm accounting: the European reporting environment, data and research perspectives. *Accounting and Business Research*, 1–45.

Biesinger, M., Bircan, Ç., & Ljungqvist, A. (2020). Value Creation in Private Equity: EBRD Working Paper No. 242. *Swedish House of Finance Research Paper*, *20*(17).

Braun, R., Jenkinson, T., & Schemmerl, C. (2016). Adverse Selection and the Performance of Private Equity Co-Investments. *SSRN Journal*.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohn, J. B., Hotchkiss, E. S., & Towery, E. M. (2022). Sources of Value Creation in Private Equity Buyouts of Private Firms. *Review of Finance*, *26*(2), 257–285.

Dalgraad, P. (2008). *Introductory Statistics with R*. Springer New York.

Engel, N., Braun, R., & Achleitner, A.-K. (2012). Leverage and performance of buyouts: (How) does the use of debt Impact equity returns? *Zeitschrift für Betriebswirtschaft*, *82*, 451–490.

European Commission. (2022). *Exchange Rate (InforEuro)*. https://ec.europa.eu/info/funding-tenders/procedures-guidelines-tenders/information-contractors-and-beneficiaries/exchange-rate-inforeuro_en. (Accessed: 22.09.2022)

Fang, L., Ivashina, V., & Lerner, J. (2015). The disintermediation of financial markets: Direct investing in private equity. *Journal of Financial Economics*, *116*(1), 160–178.

Ferman, B. (2021). Matching estimators with few treated and many control observations. *Journal of Econometrics*, *225*(2), 295–307.

French, K. R. (n.d.). *Detail for 5 Industry Portfolios.* https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_5_ind_port.html. (Accessed: 28.09.2022)

Gompers, P., Mukharlyamov, S., & Mukharlyamov, V. (2015). What Private Equity Investors Think They Do for the Companies They Buy. *Harvard Business Review*.

Graf, C., Kaserer, C., & Schmidt, D. (2009). Private Equity in Financial Institutions: Value Creation and Performance. *SSRN Journal*.

Guo, S., Hotchkiss, E. S., & Song, W. (2009). Do Buyouts (Still) Create Value? *Journal of Finance, Forthcoming*.

Hahn, M. (2009). *Essays on Private Equity Value Creation: Financial and Operating Performance in Private Equity Owned Companies* (Dissertation). Ludwig-Maximilians-Universität München.

Harris, R., Jenkinson, T., & Kaplan, S. (2013). *Private Equity Performance: What Do We Know?* Cambridge, MA.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, *47*(1), 153.

Jensen, M. C. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *The American Economic Review*, *76*(2), 323–329.

Johan, S., & Zhang, M. (2021). Information Asymmetries in Private Equity: Reporting Frequency, Endowments, and Governance. *Journal of Business Ethics*, *174*(1), 199–220.

Johnston, R., Jones, K., & Manley, D. (2018). Confounding and collinearity in regression analysis: a cautionary tale and an alternative procedure, illustrated by studies of British voting behaviour. *Quality & quantity*, *52*(4), 1957–1976.

JPMorgan. (2021). Food Fight: An update on private equity performance vs public equity markets.

Kaplan, S. (1989a). The Effects of Management Buyouts on Operating Performance and Value. *Journal of Financial Economics*, *24*.

Kaplan, S. (1989b). Management Buyouts: Evidence on Taxes as a Source of Value. *The Journal of Finance*, *44*(3), 611–632.

Kaplan, S., & Strömberg, P. (2009). Leveraged Buyouts and Private Equity. *Journal of Economic Perspectives*, *23*(1), 121–146.

Kaplan, S. N., & Schoar, A. (2005). Private Equity Performance: Returns, Persistence, and Capital Flows. *The Journal of Finance*, *60*(4), 1791–1823.

Knitterscheidt, K., & Murphy, M. (2020). *Der Notverkauf - Thyssen-Krupp verkauft Aufzugssparte für 17 Milliarden Euro: Thyssen-Krupp verkauft seine Aufzugssparte am Finanzinvestoren. Mit der Abgabe des letzten verbleibenden Gewinnbringers will der Konzern sich retten. Handelsblatt.* https://www.handelsblatt.com/unternehmen/industrie/elevator-geht-an-investoren-der-notverkauf-thyssen-krupp-verkauft-aufzugsparte-fuer-17-milliarden-euro/25587478.html. (Accessed: 19.09.2022)

Levene, H. H. (1960). Robust tests for equality in variances. In I. Olkin & H. Hotelling (Eds.), *Contributions to Probability and Statistics: Essay in Honor of Harold Hotelling* (pp. 278–292). Stanford University Press.

Ljungqvist, A., Richardson, M., & Wolfenzon, D. (2008). The Investment Behavior of Buyout Funds: Theory and Evidence. *Cambridge, MA*.

Lowenstein, L. (1985). Management Buyouts. *Columbia Law Review*, *85*(4), 730–784.

Lu, X., & White, H. (2014). Robustness checks and robustness tests in applied economics. *Journal of Econometrics*, *178*, 194–206.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *BIOMETRIKA*, *78*(3), 691–692.

Phalippou, L., & Zollo, M. (2005). What Drives Private Equity Fund Performance? *Working Paper*.

PricewaterhouseCoopers GmbH Wirtschaftsprüfungsgesellschaft. (2020). Private Equity Trend Report 2020 - Bull or bear? *With the support of Steve Roberts and Elena Naydenova*.

Puche, B. (2016). *Essays on Value Creation and its Determinants in Private Equity* (Unpublished doctoral dissertation). Technische Universität München, München. TUM School of Management.

Rikato, M. (2014). *Essays on Private Equity: Operating Performance, Investment Selection Success and Costliness of Placement Agents* (Unpublished doctoral dissertation). University of Ljubljana, Ljubljana. Faculty of Economics.

Rosenbaum, P. R. (2010). *Observational Studies* (2nd ed.). Springer.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician*, *39*(1), 33–38.

Salciccioli, J. D., Crutain, Y., Komorowski, M., & Marshall, D. C. (1973). Sensitivity Analysis and Model Validation. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 263–271). New York.

Sender, H. (2013). *Lunch with the FT: David rubenstein. the co-founder of the carlyle group is worth an estimated $3bn. over lobster bisque, he talks about his plans to give away the bulk of his fortune. financial times.* https://www.ft.com/content/97a35d8c-90ac-11e2-a456-00144feabdc0. (Accessed: 09.11.2022)

Sensoy, B. A., Wang, Y., & Weisbach, M. S. (2014). Limited partner performance and the maturing of the private equity industry. *Journal of Financial Economics*, *112*(3), 320–343.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *BIOMETRIKA*, *52*(3-4), 591–611.

Stafford, E. (2015). Replicating Private Equity with Value Investing, Homemade Leverage, and Hold-to-Maturity Accounting. *SSRN Journal*.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, *25*(1), 1–21.
Student. (1908). The Probable Error of a Mean. *BIOMETRIKA*, *6*(1), 1–25.
Wickham, H. (2014). Tidy Data. *Journal of Statistical Software*, *59*(10), 103–138.

Junior Management Science 8(3) (2023) 658-681

Junior Management Science

journal homepage: www.jums.academy

Strategic Foresight Capability and its Impact on Firm Performance: A systematic, AI-based Literature Review

Stefan Gerhard Taferner

Universität St. Gallen

Abstract

Strategic foresight is a growing field that attracts scholars aiming to reduce the uncertainty of volatile business environments. However, the field must address crucial challenges to advance theory and practice. To achieve this, the thesis presents a systematic, AI-based literature review that structures the foresight field, displays the status quo, and offers research trajectories. A sample of 243 journal-published articles is analyzed to create an organizing framework as well as provide narrative syntheses on foresight capability and its impact on firm performance. This analysis points out that foresight research often lacks theoretical foundations, mixes epistemological dimensions, and does not work toward a shared objective. Still, six research themes and their connections were identified for an organizing framework. Further, the review points out capabilities for successful foresight: Distinct processual and contextual capabilities developed in accordance with a firm's environment can ensure success. Lastly, the paper emphasizes that "successful foresight" manifests in practice through a positive impact on strategic, organizational, and performance outcomes. Those findings support the efforts of establishing foresight in management studies and improving academic progress.

Keywords: Strategic foresight; Firm performance; Foresight capabilities; Literature review; Organizing framework.

1. Introduction

In a global business environment with high uncertainty and disruptive innovations, firms find themselves in a substantially unstable competitive landscape (Tapinos & Pyper, 2018). This requires them to develop capabilities and logics that allow them to succeed despite the faced complexity (Haarhaus & Liening, 2020). This directed the attention of researchers and professionals to the idea of reducing the presently faced uncertainty by understanding possible future scenarios and their consequences (Iden, Methlie, & Christensen, 2017). The concept of "strategic foresight" captures this idea as a practice of using information about the future to systematically learn, improve decision-making, and gain a competitive advantage (Rohrbeck, Battistella, & Huizingh, 2015). Put another way, strategic foresight aims at understanding trends and changes before the competition and capitalizing on this knowledge through superior performance (Yoon, Kim, Vonortas, & Han, 2018). The interest in this idea resulted in a rapidly growing field of management research (Rohrbeck, Thom, & Arnold, 2015).

However, while strategic foresight became a prevalent topic in research (Burt & Nair, 2020), several publications point out issues in the still nascent field (e.g., Piirainen & Gonzalez, 2015). The main concerns are that the field is weakly organized, unstructured, and insufficiently linked to relevant debates in management journals (Rohrbeck, Battistella, & Huizingh, 2015). This hampers theoretical progress and the advancement of foresight practices since the field lacks a clear, combined objective (Hines, 2020; Iden et al., 2017; Snyder, 2019). Further, research to date focused mainly on antecedents and foundations for foresight while findings on necessary capabilities and outcomes are scarce despite their perceived importance (Iden et al., 2017). This paper addresses these issues through a systematic, AI-based literature review that creates orientation in the developing field and bridges the existing gaps to improve future research. In doing so, the paper answers the following three questions: Q1: "What is the current state of strategic foresight research?", Q2: "Which capabilities do firms need for successful strategic foresight?", and Q3: "How does successful strategic foresight reflect itself in overall firm performance?".

DOI: https://doi.org/10.5282/jums/v8i3pp658-681

These three questions will be answered through a systematic literature review. This review consists of three distinct parts that are interconnected but focus on one individual question each. First, a quantitative overview of strategic foresight research is provided to structure the field, identify gaps and opportunities, and ultimately answer the first research question (Q1). Then, two qualitative reviews are conducted, that focus on the current knowledge on necessary capabilities for successful strategic foresight (Q2) and foresight's impact on firm performance (Q3). In sum, this will advance the foresight field by adding the lacking structure, displaying the status quo on important frontiers, and providing guidance for future research and practice.

The data for this review is gathered primarily by the AI-based search engines Iris.ai and Semantic Scholar and will be cross-checked with the established databases ScienceDirect, JSTOR and Sage Journals to ensure exhaustiveness and to assess the maturity of the AI search engines. For the quantitative component of the review, 243 journal-published articles were compiled and nominally categorized to point out trends and themes in the field. For the two qualitative components, the findings of relevant papers are presented through narrative syntheses that create well-founded bases for further research. First, foresight capability, which research suggests as a mediator for foresight's success and outcomes, is analyzed regarding the capabilities that compose it in different conceptual models/frameworks. Second, the possible impact that successful strategic foresight can have on firm performance is displayed through collected empirical findings. The insights of those three components are contextualized in the final discussion.

Following this introduction (1) the next chapter further defines the concept of strategic foresight and displays the evolution of the research field with its gaps and challenges (2). Then, the research approach and methodology of this paper are explained in detail (3) as a basis for the following chapters. In those chapters, a tripartite systematical literature review is conducted and the findings of each part are presented. First, a quantitative review displays the existing foresight research and creates structure in the field through a comprehensive framework (4.1). Then, the existing findings on strategic foresight capability (4.2) and foresight's impact on firm performance (4.3) are showcased thoroughly in qualitative reviews. The information presented in those three literature review components create the basis to answer the paper's research questions (Q1-3) which will be done in the subsequent discussion (5). To round the paper off, the most important findings, contributions, and limitations of the paper are concluded (6).

2. Theoretical Background

This chapter discusses the concept of strategic foresight as well as its existing research and will function as a foundation and context for the thematical literature review in the main body of the paper. The following sections do not aim to develop a general theoretical model for strategic foresight but rather provide the necessary information to understand the subsequent analyses, frame the discussion, and implicitly point out the academic and practical relevance of the paper.

2.1. Conceptual Definition of Strategic Foresight

Futurism captures the idea of studying the future, learning from it, and integrating the acquired knowledge into present-day decision-making (Burns, 2021). This concept emerged due to the velocity, uncertainty, and complexity of environmental changes of our time which lowered scholars' and professionals' confidence in the effectiveness of decisions based solely on past data (Bennett & Lemoine, 2014; Hobday et al., 2020). Instead, they started to integrate evaluations of trends and environmental discontinuities into their decision-making in an effort to reduce the faced complexity (Rohrbeck & Kum, 2018). One especially prominent and increasingly researched example of such futures studies is "foresight" (Ballandonne, 2020).

Fundamentally, the concept of foresight builds on the assumption that while there are multiple possible futures, drivers of change can be identified, studied, and used to influence the future (Berger, de Bourbon Busset, & Massé, 2007). This epistemologically differentiates foresight from forecasting, which rather tries to predict one, scientifically justifiable future (Martin, 2010). "Strategic" foresight connects this idea to corporate organizations and describes a firm-level process of "identifying, observing, and interpreting factors that induce change, determining possible organization-specific implications, and triggering appropriate organizational responses" (Rohrbeck, Battistella, & Huizingh, 2015). In practice, firms facilitate this with the aim of understanding change before the competition to proactively shape their behavior and achieve better firm performance (Yoon et al., 2018).

2.2. Evolution of the Strategic Foresight Field

Despite some challenges, the study of (strategic) foresight has a long tradition and is constantly growing and evolving (Hines, 2020). Over time, the number of yearly publications steadily increased while the focus of researchers shifted significantly (Gordon, Ramic, Rohrbeck, & Spaniol, 2020). Reviews about the evolution of the field (e.g., Gordon et al., 2020; Hines, 2020; Iden et al., 2017) point out that historical research mostly focused on "methods applied, organizing practices, and experiences gained" (Iden et al., 2017) while current research concentrates on corporate integration, and foresight's impact on competitive and innovation capabilities (Gordon et al., 2020).

In particular, many recent studies focused on how successful foresight can grant a competitive advantage and how that affects firm performance (e.g., Arokodare & Asikhia, 2021; Rohrbeck & Kum, 2018). Others covered the antecedents, challenges, and opportunities of firms trying to implement strategic foresight practices in their organization (e.g., Hamel, Ims, & Yoccoz, 2022; Mastio & Dovey, 2021; Wright, O'Brien, Meadows, Tapinos, & Pyper, 2020). Further,

a new stream of foresight research emerged which focuses on how technology can be leveraged to improve strategic foresight and its outcomes (e.g., Mühlroth & Grottke, 2018; Schoemaker & Tetlock, 2017). However, many researchers argue that the theoretical progress addressing these topics is slow due to the field's weak organization and structure (Hines, 2020; Iden et al., 2017). Therefore, a detailed research overview will be presented in this paper to resolve this issue and refine future research output.

2.3. Academic Challenges in Foresight Research

As indicated, the foresight field faces some crucial challenges due to its developing state (Rohrbeck, Battistella, & Huizingh, 2015). Firstly, there is no single, generally accepted "theory of in within foresight" which results in a missing theoretical basis (Piirainen & Gonzalez, 2015). While several publications addressed this issue by providing coherent conceptual definitions (e.g., van der Laan, 2021; Gordon et al., 2020; Rohrbeck, Battistella, & Huizingh, 2015), many new studies still differ in their theoretical foundation. Secondly, confusion arises from the fact that various terms in the foresight field are used for related, overlapping concepts: Specifically, "strategic foresight", "corporate foresight", "managerial foresight", and "organizational foresight" are often used synonymously while some scholars argue that differentiation is necessary (Rohrbeck, Battistella, & Huizingh, 2015). Thirdly, building on other scholars' findings can be difficult because researchers discuss strategic foresight in different dimensionalities: Some see it as an individual phenomenon, while others describe it on an organizational level (Sarpong, Maclean, & Davies, 2013). Those factors combined cultivate an opaque research field with slow theoretical progress (Rohrbeck, Battistella, & Huizingh, 2015).

The last issue makes it especially difficult to build a shared understanding of foresight due to competing epistemologies (Paliokaitė, Pačėsa, & Sarpong, 2014; Sarpong, Maclean, & Alexander, 2013). While some researchers suggest that organizational and individual foresight processes could be considered as isomorphic (Hines, Gary, Daheim, & van Der Laan, 2017), others emphasize that differentiation is necessary (Rohrbeck, Battistella, & Huizingh, 2015). Also, it is unclear how individual foresight capability translates to an organizational level and how this effects overall firm performance. To prevent this uncertainty, this paper focuses mainly on organization-level foresight as the unit of analysis. However, individual-level foresight is covered in the organizing framework (Ch. 4.1.3) and discussion of foresight capability (Ch. 4.2.2) due to its potential role as a micro-foundation for foresight outcomes.

Also, despite significant contributions to foresight methodology and organization, other frontiers that are highly relevant to advance academia and practice have not sufficiently been tackled (Gordon et al., 2020). In particular, findings on foresight capability and foresight's advantages for firms are still relatively limited despite their practical importance (Gordon et al., 2020; Hines, 2020). Therefore, this paper does not only provide a recent, structured overview of the field but also displays the current knowledge on the two mentioned topics to showcase the theoretical status quo in those areas. Those two objectives combined aim to create a well-founded, state-of-the-art foundation and reference point for future strategic foresight research.

3. Research Approach and Methodology

The following sections explain the scientific approach of this paper and point out how the process of gathering and analyzing data was conducted. Further, the differentiation and interconnection between the quantitative and qualitative components of the literature review are displayed. This aims to build an understanding of the paper's research process as well as to provide transparency on its findings, their scientific basis, and potential limitations. Figure 1 illustrates the overall approach that will be discussed in more detail in the following sections.

3.1. Approach and Source Selection

The scientific approach in this paper is twofold and differs between the quantitative and qualitative sections: First, a quantitative review of foresight literature systematically and statistically displays the existing research to provide structure and orientation in the field. The aim here is to identify and classify existing publications to point out opportunities for future research. This part aims to answer the first research question (Q1) and functions as a foundation for the subsequent qualitative review. There, a systematic literature review focusing on existing research regarding strategic foresight capability and (successful) foresight's impact on firm performance is conducted. This review displays and summarizes the theoretical progress in those subfields to ultimately answer the second and third research question (Q2-3) of this paper.

Regarding source selection, the quantitative review includes a broad spectrum of strategic foresight research and does not set a content-related focus. It covers all English, journal-published, and peer-reviewed articles that the used search engines identify. Other sources like books or conference papers are omitted to minimize quality concerns and keep the sample in a manageable size. This does not entail a significant risk of excluding important scholarly contributions because researchers typically publish their work in academic journals first. The review will focus mainly on papers published in and after the year 2000 because environmental uncertainty and complexity increased significantly during that time which spiked the interest in business-related (strategic) foresight (Gordon et al., 2020). Prior research focused mainly on quantitative forecasting in less volatile environments (Rohrbeck, Battistella, & Huizingh, 2015) and is, therefore, less relevant for this paper (Djuricic & Bootz, 2019). The search keywords are "strategic foresight" as well as its previously introduced related "synonyms" (chapter 2.3) and are intentionally formulated broadly to cover a large spread of academic literature. However, articles that describe

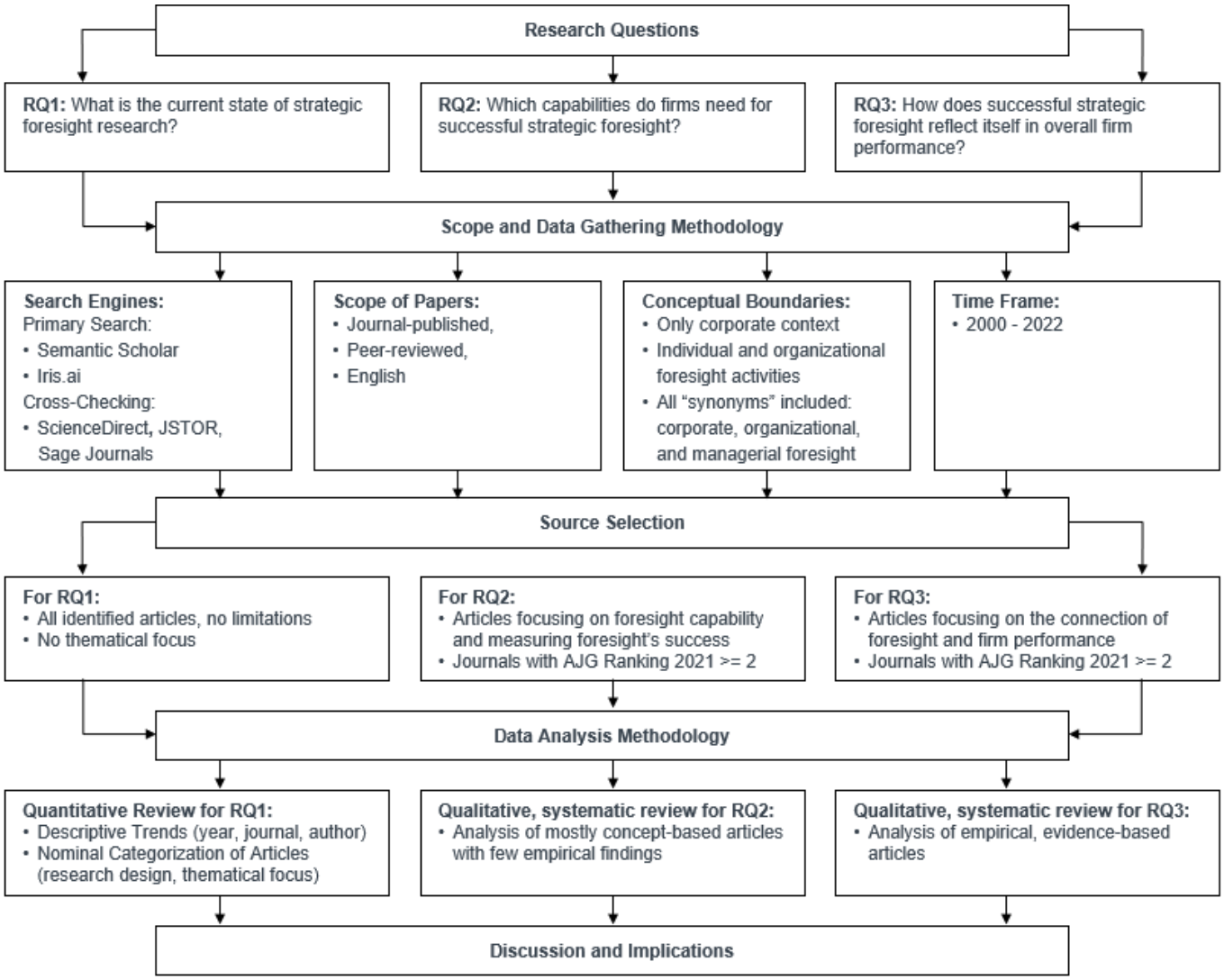


Figure 1: Research Approach (Own Illustration).

foresight that does not occur in corporate organizations (e.g., governmental foresight) will be excluded.

For the qualitative literature review, most criteria for in- and exclusion are the same as in the quantitative part. However, the thematical focus is narrower and only covers the stated topics (capability and firm performance) while unrelated foresight research will be excluded. Additionally, the quality of the analyzed articles is relevant in this part and only well-respected research will be integrated into the final discussion (AJG Ranking 2021: >= 2). This differentiates the qualitative literature review from the preceding quantitative review that includes all journal-published articles (regardless of their rating and reputation). In the chapter about foresight capability (Ch. 4.2), an exception is made for two book-published models that conceptualize foresight capability (Miller & Sandford, 2019; Rohrbeck, 2010) since they are frequently referenced but never fully explained in reputable journals. This is necessary due to a lack of specific, evidence-based findings on strategic foresight capability (e.g., because of confidentiality agreements) which results in a small knowledge base overall (Daheim & Uerz, 2008).

3.2. Data Gathering Methodology

The data in this paper is primarily gathered by the AI-based research engines "Iris.ai" and "Semantic Scholar" but cross-checked with the traditional databases Elsevier (Science Direct), JSTOR (Journal Storage), and SAGE Journals. The reason for this selection, the potential benefits of the AI-based engines, the motivation behind the cross-checking approach, and an explanation of which data will be extracted for the literature review are described below.

In short, the two primarily used search engines use artificial intelligence to understand the semantics of scientific literature to improve the research process and its findings (Extance, 2018). In theory, the tools offer the potential to increase the breadth of data and make the findings more reproducible while enabling a faster data gathering process (Extance, 2018). Iris.ai does that by receiving a paper or a

problem statement as an input and then "fingerprinting" the information based on extracted keywords, contextual synonyms, and hypernyms (Iris.ai, 2022). This fingerprint is then matched against >200M papers to create an "explore map" of connected papers that can be narrowed down into a precise reading list (Iris.ai, 2022). Semantic Scholar resembles traditional search tools but provides additional, more focused information (Extance, 2018). It advances search through capturing popularity metrics, indirect citations, data sets, methods, and connections of relevant articles (Extance, 2018). It uses NLP to extract information from papers to build a reading list that can be adapted iteratively (Extance, 2018).

In addition, Elsevier, JSTOR, and SAGE Journals are used as secondary sources to find relevant research for two main reasons: Firstly, the two AI research engines are not fully established yet which makes their exhaustiveness questionable. Therefore, those databases that cover "the vast majority" of foresight literature (Iden et al., 2017; Marinković, Al-Tabbaa, Khan, & Wu, 2022) are used to ensure that the literature review in this paper provides a comprehensive overview. Secondly, the two AI engines are used in an effort to assess the current maturity and convenience of such search tools as a methodological innovation. Cross-checking the covered literature with the traditionally recognized databases will allow conclusions on this matter (see chapter 5.3).

The type of data gathered for the literature review differs between the quantitative and qualitative components: For the quantitative review, the title, author/s, publishing date, thematical focus, research design, and theoretical basis of each identified paper are extracted. This information is collected in an excel sheet and functions as the basis for a statistical display (chapter 4.1). For the two qualitative components, the findings of relevant papers covering the observed topics are extracted, summarized, and contrasted to build well-founded answers to the research questions. Here, the data is not used in isolation but in the context of its respective study.

3.3. Data Analysis Methodology

The analysis of the gathered data will be done manually, due to the engine's limitations in that aspect. Here, it must again be distinguished between the methodology for the quantitative and qualitative components of the literature review. The following paragraphs explain how the analyses for each part are conducted and how the findings of this process are synthesized.

Firstly, the analysis in the quantitative review will be a statistical assessment that aims to showcase trends and gaps in the existing literature based on a nominal categorization of identified articles. More precisely, the extracted data is used to display how the number of yearly publications changed over time as well as how those publications are distributed among different journals and researchers. Further, the articles are categorized according to their research design and thematical focus to point out what researchers have historically focused on and how they achieved their findings. The categorization of the research design follows the schema of Orlikowski and Baroudi (1991): The contributions are grouped into either conceptual (frameworks, models, reviews) or empirical designs (surveys, interviews, case studies, experiments). The different thematical foci were established by reading (parts of) the articles in the sample and split into antecedents, foresight capability, organizational foundations, individual micro-foundations, moderators, and outcomes of foresight. The analysis of this categorization is done in Excel and displayed graphically in the text while the raw data is shown in Appendix A. In addition to its objective of organizing the field, this analysis functions as the basis for the succeeding qualitative analyses since the categorization filters out papers on capabilities for successful strategic foresight (capability) and its impact on firm performance (firm performance).

Those papers are then analyzed in more detail and synthesized in the qualitative, systematic literature reviews. The qualitative analysis of the papers' findings is done deductively by reading text elements word by word. Then, narrative syntheses are created due to their suitability to create comprehensive overviews of heterogenous fields (Marinković et al., 2022). This enables a display of the combined research findings in a "storytelling-manner" (Bailey, 2006) that is continued iteratively until well-founded answers to the two research questions (Q2-3) are found. Since empirical research on necessary capabilities for successful strategic foresight is limited, the analysis builds on conceptual frameworks rather than practical evidence: Competency / Capability models that suggest components of foresight capability and measure the level of individual or organizational foresight capability are introduced and compared to draw conclusions on overarching strategic foresight capability. If those models were already tested empirically on their connection to firm performance, those results are also presented. Next, the paper provides an analysis regarding the findings on strategic foresight's impact on firm performance. Here, firm performance is not confined to external results but also includes internal outcomes like improved innovation capabilities that do not yield instant (external) returns. The objective of this analysis is to build a theoretical basis for future research on these frontiers as well as to contextualize the findings within the organizing framework developed in chapter 4.1.3. However, the paper does not aspire to provide a complete overview that includes all studies but rather aims to incorporate respected ideas and defining trends. While the thematical literature review presents those ideas, the subsequent discussion will put them into perspective.

4. Themes in Strategic Foresight Research

This systematic literature review is split into three different components: First, the current state of strategic foresight research is displayed quantitatively to point out trends, display gaps, and create an organizing framework of the field. Second, qualitative findings on necessary capabilities for successful foresight as well as moderators for foresight's success

are explained. Third, findings on successful foresight's impact on firm performance are compiled in the final section.

4.1. Current State of Strategic Foresight Research

In this first part of the review, the existing literature on strategic foresight is displayed in an effort to provide more transparency in the developing field. In the first two sections (4.1.1 and 4.1.2), the goal is to quantitatively show the existing research and point out research streams. In the third section (4.1.3) the aim is to contextualize this existing literature in an organizing model to showcase links, gaps, and trajectories. In combination, the sections provide the basis to answer research question Q1 ("What is the current state of strategic foresight research?").

4.1.1. Descriptive Trends in Strategic Foresight Research

Similar to preceding reviews (e.g., Iden et al., 2017; Singh, Dhir, Das, & Sharma, 2020), this study finds that the number of yearly publications has been steadily increasing since environmental uncertainty created interest in strategic foresight (Gordon et al., 2020). Overall, this literature review compiles 243 journal-published articles on strategic foresight and finds a CAGR of 9% in yearly publications between 2000-2022. Of those articles, only 18% were published before 2010 while around 57% of all papers were published from 2015 to 2000. The year with the most yearly publications is 2020 with 26 articles (11%) followed by 25 articles in 2015 (10%). When combined, the timeframe 2010-2022 constitutes 82% of all publications on strategic foresight which is congruent with other studies that point out the growing scientific importance of the topic (e.g., Marinković et al., 2022; Rohrbeck, Battistella, & Huizingh, 2015). Figure 2 illustrates this development.

Those publications stem from a variety of different publishers ranging from business- and management reviews to technology journals. In this paper's sample, around a third of all identified articles (n=80) were published in the journal Technological Forecasting and Social Change (33%). Following that, around 19% were published in Futures (n=46) and 8% in Technology Analysis & Strategic Management (n=20). Those three journals historically offer the highest research output on strategic foresight and constitute approximately 60% of all articles in the sample. The remaining publications span a total of 66 different journals with various foci. Table 1 briefly illustrates the literature's split between some of the journals while the full breakdown can be found in Appendix A. Here, it must be noted that a significant number of publications (n=142) stems from journals with a comparatively low ranking (AJG Ranking 2021: <= 2) while only very few (n=4) were published in highly reputable (management) journals (AJG Ranking 2021: > 3). This is in line with the critique that foresight research is insufficiently linked to relevant debates in respected management journals (Rohrbeck, Battistella, & Huizingh, 2015) and again points out that the still nascent field must develop its foundations (Piirainen & Gonzalez, 2015).

The authors of those articles are numerous with a total of 419 different researchers that contributed to journal-published articles. Some of the researchers with the highest quantitative output are Rohrbeck (n=14), Sarpong (n=12), Vecchiato (n=7), and Wright (n=6). A more extensive breakdown of the different authors is displayed in Table 2. Congruent with Iden et al. (2017), this breakdown indicates that historically a dominant proportion of foresight research was conducted by European scholars. Regarding research designs, this paper uses the categorization of Orlikowski and Baroudi (1991) into conceptual and empirical designs. Here, the conceptual design groups all articles that create concepts, frameworks, or models (including literature reviews) while empirical research covers all approaches that utilize some form of empirical data (e.g., surveys, interviews, case studies, experiments, or data mining). In the sample, the most prominent research design are conceptual approaches (n=107) followed by case studies (n=88) and surveys (n=26). For mixed approaches (e.g., case study with interviews) the studies were categorized according to the "dominant" approach. The breakdown of used approaches is illustrated in Table 3. This breakdown implies that strategic foresight research mostly builds on qualitative approaches which seems logical since foresight was developed as an addition to traditionally quantitative forecasting (Marinković et al., 2022; Martin, 2010).

Overall, this display of trends in the foresight field is in line with existing reviews but provides an updated, more extensive picture. The display shows that publications on strategic foresight are becoming more frequent and span a broader range of journals and researchers. While the topic is gaining importance, its relevance for the general management discourse is still limited. Many scholars attribute this to differing or unclear theoretical foundations between studies (Piirainen & Gonzalez, 2015). Thus, some of the most relevant foundations are explained next.

4.1.2. Theoretical Research Streams in Foresight Studies

As mentioned in the introduction, researchers argue that many articles in the foresight field lack a clear theoretical foundation (Piirainen & Gonzalez, 2015). This argument is supported by the observations of this paper which found that around 35% (n=85) of the articles in the final sample do not build on a specific theoretical concept. For the remaining articles, the authors based their work on various theories with network theory, dynamic capabilities, and organizational learning as the most frequently used examples. Those theories and the frequency of their usage are illustrated in Table 4 and briefly put into context in the following paragraphs.

Firstly, dynamic capabilities theory explains how organizations can ensure competitiveness through developing certain organizational capabilities (Teece, Pisano, & Shuen, 1997). This theory conceptualizes foresight as a micro-foundation or antecedent for such capabilities (e.g., Haarhaus & Liening, 2020) or even suggests foresight as a distinct organizational capability in itself (e.g., Pulsiri & Vatananan-Thesenvitz, 2021; Rhisiart, Miller, & Brooks,

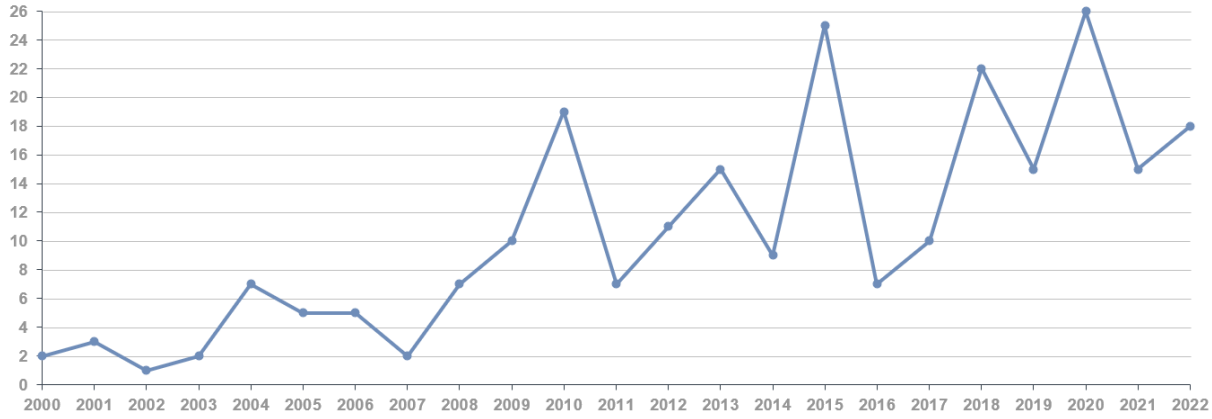


Figure 2: Yearly Publications from 2000-08/2022 (Own Illustration).

Table 1: Number of publications in different journals.

Journal	Articles	Percentage
Technological Forecasting and Social Change	80	33
Futures	46	19
Technology Analysis & Strategic Management	20	8
Foresight	11	5
World Futures Review	4	2
IEEE Transactions on Engineering Management	4	2
Journal of Futures Studies	3	1
Long Range Planning	3	1
Other Journals (number of articles per journal <= 2)	72	29
Total	243	100

Table 2: Most frequently published Researchers.

Researchers	Country	Contributions
Rohrbeck	Denmark	14
Sarpong	UK	12
Vecchiato	Italy	7
Von der Gracht	Germany	6
Wright	UK	6
Burt	UK	5
Gordon	UK	5
Others (n=412)	-	466

Table 3: Used Research Designs.

Research Methods	Articles	Percentage
Conceptual	107	44
Empirical	136	56
- Case Studies	88	36
- Surveys	26	11
- Interviews	13	5
- Experiments	5	2
- Data Mining	3	2

Table 4: Theoretical Foundations and Research Streams in the Foresight field.

Theoretical Foundation	# Articles	Percentage	Influential Studies
Dynamic Capabilities	36	15	Haarhaus & Liening, 2020; Pulsiri & Vatananan-Thesenvitz, 2021; Ramírez, Österman, & Grönquist, 2013
Network Theory	21	9	Calof, Arcos, & Sewdass, 2018, Adegbile, Sarpong, & Meissner, 2017; Heger & Boman, 2015; Nugroho & Saritas, 2009; Van der Duin, Heger, & Schlesinger, 2014; Weber, Sailer, & Katzy, 2015
Organizational Learning	45	18	Bootz, Monti, Durance, Pacini, & Chapuy, 2019; Bootz, 2010; Eskandari, Mohammadi, & Rahimi, 2020; Peterson & Wu, 2021; Pulsiri & Vatananan-Thesenvitz, 2021; Yoon et al., 2018
Others	56	23	-
No Theoretical Foundation	85	35	-

2015). Therefore, foresight is either understood as an individual phenomenon (micro-foundation) or on an organizational level (capability) which again illustrates the problem of competing epistemologies (Sarpong, Maclean, & Davies, 2013). Regardless, this theoretical foundation sees the cultivation of foresight and other organizational capabilities as a desirable outcome that can grant a competitive advantage (Vecchiato, 2015).

Secondly, network theory explains an organization and its environment as a network of relationships and views foresight as a facilitator for such relations (e.g., Heger & Boman, 2015; Nugroho & Saritas, 2009). This is a newer theoretical foundation that was initiated by more collaborative, open foresight activities in practice that contrast with earlier, less participative foresight processes (Wiener, Gattringer, & Strehl, 2020). Research on this foundation covers both, inter-personal and inter-organizational relations and is often connected to observing foresight's impact on innovation capabilities (Heger & Boman, 2015). Therefore, it tries to understand the consequences of process designs and organizational integration which suggests an organization-level analysis of foresight (Rohrbeck, Battistella, & Huizingh, 2015). In the organizing framework developed in the next chapter, it would consequently be categorized as a foundation for foresight capability.

Thirdly, organizational learning theory explains how organizations generate knowledge (through foresight) and how this knowledge is then transferred between its members (Bootz et al., 2019; Yoon et al., 2018). It covers how foresight influences an organization's learning curve and illustrates how knowledge is created and retained over time (Bootz, 2010; Peterson & Wu, 2021). Research based on this theoretical foundation often observes how foresight processes should be organized to ensure the best learnings for both organizations and individuals (Gattringer & Wiener, 2020). Organizational learning, therefore, analyzes foresight at an organizational level but also includes the individual level as a micro-foundation (Hines & Gold, 2015).

While those three theories are examples of used foundations for foresight research, more than a third of the studies (n=85) in the paper's sample do not build on one clear theoretical concept. Earlier reviews also pointed this fact out and emphasized that it is a major issue that hampers theoretical progress in the developing field (Adegbile et al., 2017; Iden et al., 2017). Therefore, creating a unified theoretical foundation is an important, unsolved issue for future research.

4.1.3. Organizing Framework of Strategic Foresight Research

The screening of articles in this quantitative review enabled the identification of theoretical foci of papers in the foresight field. This knowledge was used to develop an organizing framework for strategic foresight research which is displayed in Figure 3. The identified foci/themes are (1) antecedents of strategic foresight, (2) foresight capability, (3) organizational foundations, (4) individual micro-foundations, (5) moderators, as well as (5) foresight's impact on firm performance. The relations between those topics are displayed through arrows. Here, dotted lines suggest uncertain relations with little conducted research while solid lines show an intensively researched connection. The main quantitative (number of articles) and qualitative insights (subtopics) of the six themes are explained below while the findings on foresight capability, moderators, and firm performance are discussed in more detail in chapters 4.2 and 4.3.

This process model organizes existing research on strategic foresight and emphasizes that the different themes are interconnected. Once again, this model shows that differentiation between individual-level and organization-level foresight is necessary and that the exact relation between those dimensions is uncertain. While the model suggests that the organizational level builds on individual micro-foundations, research does not attest to how exactly the dimensions influence each other (Sarpong, Maclean, & Davies, 2013). Nevertheless, the organizing model presents a research-based perspective of the relevant relations and points out what scholars have historically focused on. Those six themes and their components are explained in the following paragraphs.

First, multiple journal-published articles focus on organizational antecedents for strategic foresight. To be exact, a total of 63 publications out of the sample (n=243) dealt with

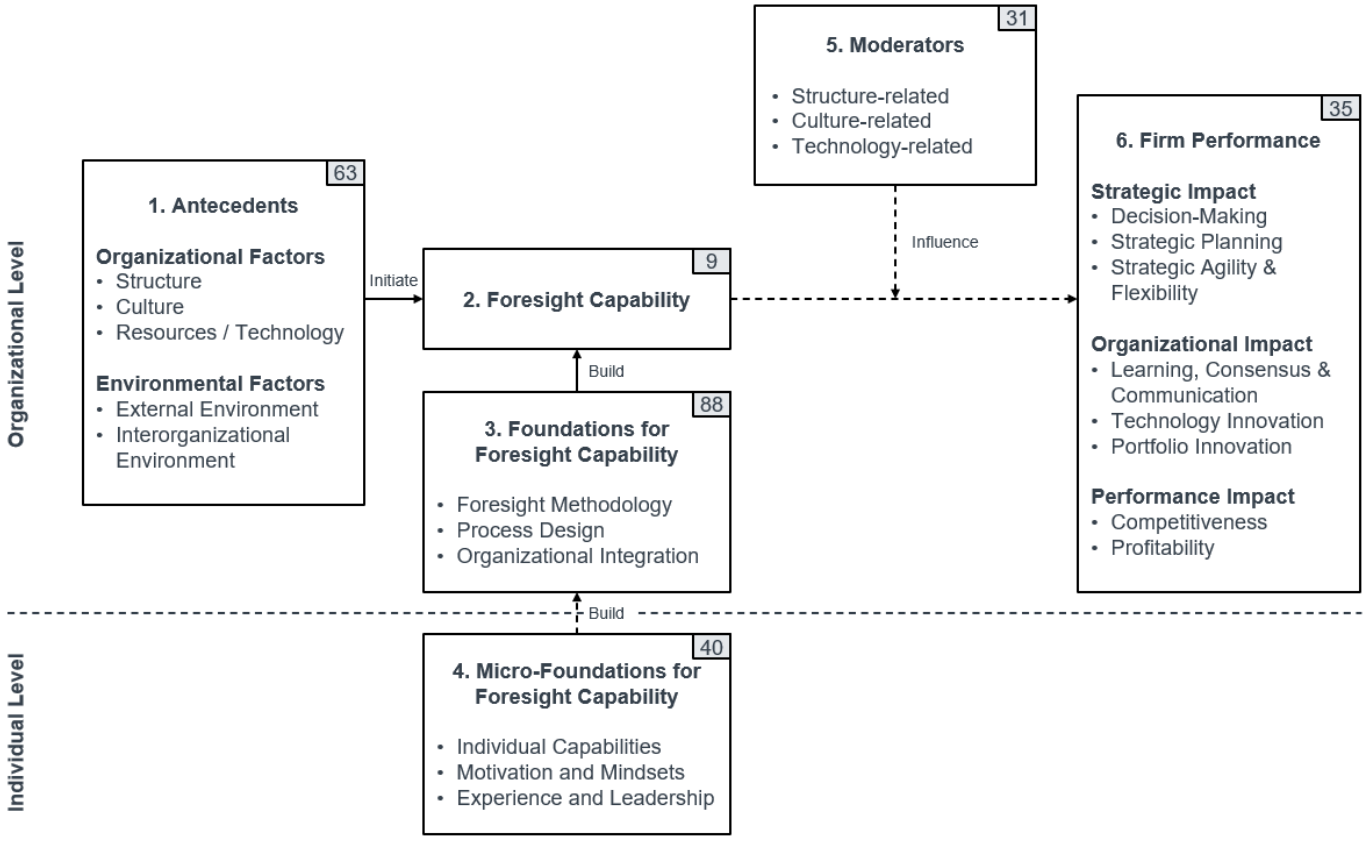


Figure 3: Organizing Framework for Strategic Foresight Research (Own Illustration).

antecedents which can be differentiated into organizational and environmental factors. Organizational factors mostly concern the structure (e.g., Ahlqvist & Kohl, 2016; Battistella, 2014), culture (e.g., Haarhaus & Liening, 2020; Wiener, Gattringer, & Strehl, 2018), and resources (e.g., Ghayoor, Rastegari, & Hosseini, 2020; Greenblott, O'Farrell, Olson, & Burchard, 2019) of an organization with foresight practices. Further, environmental antecedents are split into external (e.g., Costanzo, 2004; Vecchiato, 2012) and internal environments (Savioz & Blum, 2002; Wiener et al., 2018) and try to understand the faced uncertainty. These types of antecedents affect foresight activities (and their success) and were, therefore, sometimes not only analyzed in isolation but also in their moderating role (see chapter 4.2.4).

Second, a thematical focus in the foresight field is organizational foresight capability and its mediating effect for foresight's outcomes. In the sample, 9 articles were identified that primarily deal with foresight capability and provide frameworks that conceptualize and measure the phenomenon (e.g., Day & Schoemaker, 2005; Rohrbeck & Kum, 2018). Those articles conceptually propose what overarching capabilities are necessary to conduct foresight successfully and offer approaches to measure the maturity of those capabilities (e.g., Grim, 2009; Rohrbeck, 2010). Some articles go further and empirically investigate how the attained maturities impact outcomes of foresight activities manifested in firm performance (e.g., Rohrbeck & Kum, 2018). However, there is limited empirical evidence on the exact effect that the proposed capabilities have on firm performance in isolation and it is unclear how the (foresight) capabilities of individuals affect foresight capability at an organizational level (Daheim & Uerz, 2008). Overall, foresight capability is a theme with comparably low research output despite its perceived practical relevance (Hines, 2020). More research was conducted on foundations and micro-foundations that affect the attained capabilities. They are explained in the following two paragraphs.

Third, the organizational foundations for foresight capabilities are a prevalent research topic. In total, 88 articles in the sample with this thematical focus were identified and differentiated between foresight methodology, process design, and organizational integration of foresight. Regarding foresight methodology, research focused mainly on used tools (e.g., Godet, 2000; Idoko & MacKay, 2021), technological aids (e.g., Boe-Lillegraven & Monterde, 2015; Gibson, Dime, Garces, & Dabich, 2018), and information systems that guided foresight (e.g., Arokodare, Makinde, & Fakunmoju, 2020; Von der Gracht, Bañuls, Turoff, Skulimowski, & Gordon, 2015). Studies on process design rather focused on how foresight activities are organized and differentiate between non-participative (e.g., Djuricic & Bootz, 2019; Dufva & Ahlqvist, 2015), participative (e.g., Heger & Boman, 2015; Heger & Rohrbeck, 2012), collaborative (e.g., Gattringer, Wiener, & Strehl, 2017; Weigand, Flanagan, Dye, & Jones,

2014), and network foresight processes (e.g., Nugroho & Saritas, 2009; Van der Duin et al., 2014). Finally, studies investigated how foresight is institutionalized and integrated into an organization (e.g., Paliokaitė et al., 2014). Those different factors are believed to directly affect firms' foresight capability and indirectly impact the achieved final outcomes as moderators (Rohrbeck & Gemünden, 2011).

Fourth, a theme in foresight research are the individual micro-foundations for organizational foresight capabilities. Here, a total of 40 articles in the sample were identified that cover individual foresight capabilities as well as other personal factors like motivations, mindsets, experiences, and leadership in foresight activities. Articles on individual foresight capability assess which capabilities individuals require to be "good futurists" and sometimes propose models to measure the maturity of those individual capabilities or "overall capability" (e.g., Hines et al., 2017; Rhisiart et al., 2015). Here, it is still unclear how those capabilities translate to organizational foresight capability when aggregated. Regarding personal factors, the most frequently researched subtopics are mindsets and motivations for foresight (e.g., Reid & Zyglidopoulos, 2004; Vecchiato & Roveda, 2010a), experiences in practice (e.g., Costanzo, 2004; De Smedt, Borch, & Fuller, 2013), as well as leadership styles and their relation to foresight (e.g., Luzinski, 2014; Moore, 2018; Reimers-Hild, 2018). Those factors are sometimes not only seen as direct micro-foundations for organizational foresight capability but also as moderators for foresight's impact on firm performance (e.g., Haarhaus & Liening, 2020; Wiener et al., 2020).

Fifth, a thematical focus in foresight research are moderators for the relation of foresight activities and their outcomes as well as the potential effects those moderators can have. In total, 31 articles in the sample were identified that study different types of moderators that can broadly be categorized as structure-related, culture-related, and technology-related moderators. Regarding structure-related moderators, prominently mentioned themes are hierarchical structures, the institutionalization of foresight activities, and their process designs (e.g., Haarhaus & Liening, 2020; Vecchiato, 2020). Culture-related aspects that are considered as moderators are shared values, managerial mindsets, and the overarching organizational culture (e.g., Wiener et al., 2020; Yoon, Kim, Vonortas, & Han, 2019). Lastly, technology-related moderators are the technologies used directly to conduct foresight activities and also information technologies that indirectly affect the process (e.g., Heger & Boman, 2015; Rohrbeck & Gemünden, 2011). Generally, research suggests that all those moderators can facilitate or inhibit positive outcomes of foresight activities (Sarpong, Maclean, & Davies, 2013). While the research output on such moderators is also rather limited, the number of publications has steadily increased in recent years.

Sixth, a focus of foresight research lies on the outcomes of foresight activities and their impact on firm performance. In the sample, a total of 35 articles with a focus on such outcomes were identified. Those articles do not only cover foresight's impact on external performance but also concern foresight's internal impact on a firm's strategy, organization, and innovation. Researched subtopics of the strategic impact are decision-making, strategic planning as well as strategic flexibility and agility (e.g., Gershman, Bredikhin, & Vishnevskiy, 2016; Haarhaus & Liening, 2020; Schwarz, Ram, & Rohrbeck, 2019). The research on organizational outcomes focused mainly on learning, communication, and innovation (e.g., Paliokaitė & Pačėsa, 2015; Schweitzer, Hofmann, & Meinheit, 2019; Wiener et al., 2020). Lastly, the impact on external performance was researched less frequently but some studies do examine foresight's impact on competitiveness and profitability (e.g., Arokodare & Asikhia, 2021; Boe-Lillegraven & Monterde, 2015; Rohrbeck & Kum, 2018).

Overall, those six themes organize the current knowledge on strategic foresight and the mediating role of foresight capabilities for organizational performance. Since the theoretical progress on foresight capabilities, moderators, and firm performance is rather slow while the other themes have already intensively been researched, the following chapters will explain the state of the three mentioned themes in more detail. This is done in an effort to provide a solid foundation for future research to accelerate research output. The aim here is to display driving ideas and concepts that scholars can use in their (empirical) studies to advance research and practice.

4.2. Necessary Capabilities for Successful Strategic Foresight

Research on necessary capabilities for successful strategic foresight is rather limited and opaque. However, some studies conceptualize foresight's success and provide practical implications through developing conceptual models that suggest necessary capabilities and their measurement. These models, categorized into frameworks for (1) individual and (2) organizational foresight capability, are explained and contrasted in the following paragraphs. Since research (implicitly) suggests foresight capability as a mediator for foresight's outcomes, this chapter is rounded off with insights on relevant moderators that also affect this relationship. Through this display, the chapter aims to provide the basis to answer the second research question Q2 of this paper ("Which capabilities do firms need for successful strategic foresight?").

4.2.1. Conceptualization of Successful Foresight

In order to discuss necessary capabilities for successful strategic foresight, it must first be established what success in the foresight field consists of. While there is no clear definition or measurement of successful foresight (Amsteus, 2008), there are articles that provide conceptual descriptions and suggest different components of success in foresight. The following paragraphs describe those components before the measurement of overall success in foresight activities is explained. This poses as the foundation for the discussion of foresight capability.

Table 5: Foresight Capability Research Structure.

Research Theme	Topic	Influential Studies
Conceptualization of Successful Foresight	I. Components of Successful Strategic Foresight	Aichouni, Touahmia, Kolsi, Alghamdi, & Al-Homaid, 2021; Iden et al., 2017; Maertins, 2016; Aichouni et al., 2021; van der Laan & Erwee, 2012; Bezold, 2010; Rasmussen, Andersen, & Borch, 2010; Amsteus, 2008; Wright, Van der Heijden, Burt, Bradfield, & Cairns, 2008
	II. Measuring Foresight's Success	Rincón & Díaz-Domínguez, 2022; Rohrbeck & Kum, 2018; Boe-Lillegraven & Monterde, 2015; Vecchiato, 2012; Amsteus, 2008
Individual Strategic Foresight Capability	I. Foresight Competency Model (FCM)	Hines et al., 2017; Rincón & Díaz-Domínguez, 2022
	II. Futures Literacy Framework (FLF)	Miller, 2018; Miller & Sandford, 2019; Hines et al., 2017; Rincón & Díaz-Domínguez, 2022
Organizational Strategic Foresight Capability	I. Peripheral Vision Capabilities	Day & Schoemaker, 2005
	II. Foresight Maturity Model (FMM)	Grim, 2009; Day & Schoemaker, 2005
	III. Maturity Model of Corporate Foresight	Rohrbeck, 2010; Grim, 2009; Day & Schoemaker, 2005
	IV. Future Preparedness Model (FP)	Rohrbeck & Kum, 2018; Rohrbeck, 2010; Grim, 2009; Day & Schoemaker, 2005
Moderators for Strategic Foresight's Success	I. Structure-related Moderators	Vecchiato, 2020; Haarhaus & Liening, 2020; Rohrbeck & Kum, 2018; Peter & Jarratt, 2015; Farrington, Henson, & Crews, 2012; Rohrbeck & Gemünden, 2011
	II. Culture-related Moderators	Wiener et al., 2020; Haarhaus & Liening, 2020; Yoon et al., 2019; Schwarz et al., 2019; Sarpong & Maclean, 2016; Peter & Jarratt, 2015; Daheim & Uerz, 2008
	III. Technology-related Moderators	Haarhaus & Liening, 2020; Heger & Boman, 2015; Rohrbeck, Battistella, & Huizingh, 2015; Van der Duin et al., 2014; Rohrbeck & Gemünden, 2011

Research points out that the main motivations for foresight are improving decision-making, long-term planning, and innovation capabilities as well as earlier identification and reaction to environmental changes (Rasmussen et al., 2010). Therefore, many articles define success as the achievement of those objectives through employing organizational foresight (e.g., Aichouni et al., 2021; Iden et al., 2017; Maertins, 2016). Others focus on final outcomes and describe successful foresight as foresight activities that have a positive (financial) impact on organizations overall (Bezold, 2010; van der Laan & Erwee, 2012). Therefore, whether foresight is "successful" depends on subjective organizational goals and their attainment (Maertins, 2016).

Further, it is often difficult to measure whether those goals are attained due to their predominantly qualitative nature (Rasmussen et al., 2010): Success of foresight activities is often only vaguely assessed in the long term (Boe-Lillegraven & Monterde, 2015; Vecchiato, 2012). However, scholars suggest that certain capabilities can ensure that foresight activities are conducted in a way that makes success more probable either way (e.g., Hines et al., 2017; Iden et al., 2017; Rohrbeck & Kum, 2018). Therefore, this paper defines "successful foresight" as foresight that yields positive outcomes for an organization's objectives and sees "necessary capabilities" as capabilities that make such positive outcomes more probable. Conceptual frameworks, in which such capabilities were defined, are introduced in the following sections.

4.2.2. Individual Foresight Capability

As mentioned in the introduction, research either conceptualizes strategic foresight as an individual or organizational phenomenon (Rohrbeck & Kum, 2018). Therefore, foresight capability is currently assessed and measured in different dimensions by different models. While the paper mainly focuses on organization-level foresight, this section provides information on individual foresight capability to ensure exhaustiveness since this concept, on aggregate, also affects foresight's impact on firm performance (Hines et al., 2017; Rohrbeck & Kum, 2018). The two most influential models in this category, the "Foresight Competency Model" by Hines et al. (2017) and the "Futures Literacy Framework" by Miller (2018), are described below.

Hines et al. (2017) see foresight as an individual, cognitive phenomenon as opposed to an organizational activity. They developed the Foresight Competency Model (FCM) to

identify capabilities necessary for individuals to become "successful futurists" and to measure individuals' foresight capability (Hines et al., 2017). In this model, they define foresight as an innate ability to "develop images of the future" that can be improved over time (Hines et al., 2017). Foresight capability is defined as a set of distinct individual capabilities needed to successfully conduct foresight (Hines et al., 2017). Those capabilities are segmented into six "core competencies" which are supported and contextualized by three "foundational competencies" and two "professional competencies" (Hines et al., 2017). The model suggests that at the core, a futurist must be able to scope projects (framing), explore change signals (scanning), identify possible futures (futuring), commit to a preferred future (visioning), develop results based on this information (designing), and stay flexible to alternative futures (adapting) (Hines et al., 2017). In addition, futurists require academic, personal and workplace skills (foundational competencies) as well as occupation and sector knowledge (professional competencies) to fully attain foresight capability (Hines et al., 2017). This is illustrated in Figure 4.

Overall, the Foresight Competency Model by Hines et al. (2017) sees foresight capability as an ability that is attained through the development of the above-mentioned competencies. By achieving proficiency in those competencies, individuals build foresight capability which can result in positive organizational outcomes (Hines et al., 2017). Hines et al. (2017), therefore, see foresight capability as a mediator for the activity-outcome relationship of foresight and suggests distinct capabilities that individuals can improve to facilitate positive outcomes.

The Futures Literacy Framework (FLF) by Miller (2018) also describes foresight as an individual phenomenon. In this model, "futures literacy" is defined as a capability that allows individuals to deliberately "use-the-future" and adapt the usage depending on its context and goal (Miller, 2018). In contrast to the Foresight Competency Model, this framework differentiates between non-conscious anticipation and conscious, learned anticipation which constitutes futures literacy (Miller & Sandford, 2019). According to Miller (2018), this perspective reinforces the idea that foresight capability as a skill can be developed and improved over time. This skill does not only involve dealing with the future itself but also further situational decisions: Individuals must know why and how they use the future based on the specific context (Miller & Sandford, 2019). Foresight capability, in this sense, describes the ability to utilize the future for distinct goals and to flexibly adapt the approach if necessary (Miller, 2018). Based on this definition, the Futures Literacy Framework provides a range of anticipatory systems to assist individual's foresight activities in different contexts (Miller, 2018). Overall, the model describes foresight capability more descriptive and theoretical than the Foresight Competency Model and does not mention specific capabilities necessary for success. Nevertheless, it offers practical implications by emphasizing the importance of foresight's context.

In conclusion, both discussed models on individual foresight capability see it as a developable ability that consists of distinct elements necessary to conduct successful foresight. While the models cover foresight capability on an individual level, research suggests that aggregated skills of individuals could hypothetically result in organizational foresight capability and influence the activity-outcome relationship of foresight processes. However, it is uncertain how the capability translates from an individual to an organizational level. Research on organizational foresight capability is presented next before moderators for foresight's success are introduced.

4.2.3. Organizational Foresight Capability

Other scholars define foresight as an organizational phenomenon and develop models to measure and understand its capability on this overarching level (e.g., Day & Schoemaker, 2005; Rohrbeck & Kum, 2018). The four most prominent models in this category are "Peripheral Vision Capabilities" by Day and Schoemaker (2005), the "Foresight Maturity Model" (FMM) by Grim (2009), the "Maturity Model of Corporate Foresight" by Rohrbeck (2010), and the "Future Preparedness Model" (FP) by Rohrbeck and Kum (2018). They conceptualize organizational foresight capability and offer practical implications for its development and outcomes. The following paragraphs introduce those models to draw conclusions on necessary capabilities.

The first model on this matter, developed by Day and Shoemaker (2005), describes organizational foresight capability as "Peripheral Vision Capabilities". Foresight capability in the model consists of two distinct elements: capability and need for peripheral vision (Day & Schoemaker, 2005). With this differentiation, the scholars emphasize that capability results from matching ability with need: While complex, volatile environments require high peripheral vision, stable environments have lower requirements (Day & Schoemaker, 2005). Further, "too much" peripheral vision for the specific environment can even be a disadvantage due to resulting neuroticism and inefficiencies (Day & Schoemaker, 2005; Rohrbeck & Kum, 2018). Day and Schoemaker (2005) propose that organizations must understand their need for peripheral vision by assessing the nature of their strategy as well as the volatility and complexity of their surroundings. Then, the organizational capability for peripheral vision can be assessed to determine the (relative) foresight capability (Day & Schoemaker, 2005). Peripheral vision capability in this model consists of five contextual elements: leadership orientation, strategy making, knowledge management, organizational configuration, and culture (Day & Schoemaker, 2005; Rohrbeck & Kum, 2018). The combination of those elements is illustrated in Figure 5.

In practice, the maturity of those elements is quantified through management surveys and compared to the organization's need for peripheral vision. This comparison results in an assessment of foresight capability and offers implications on areas to improve (Day & Schoemaker, 2005). Overall, an organization's match of peripheral vision capabilities needed and attained is assumed to determine foresight's effect on firm performance (Day & Schoemaker, 2005). There-

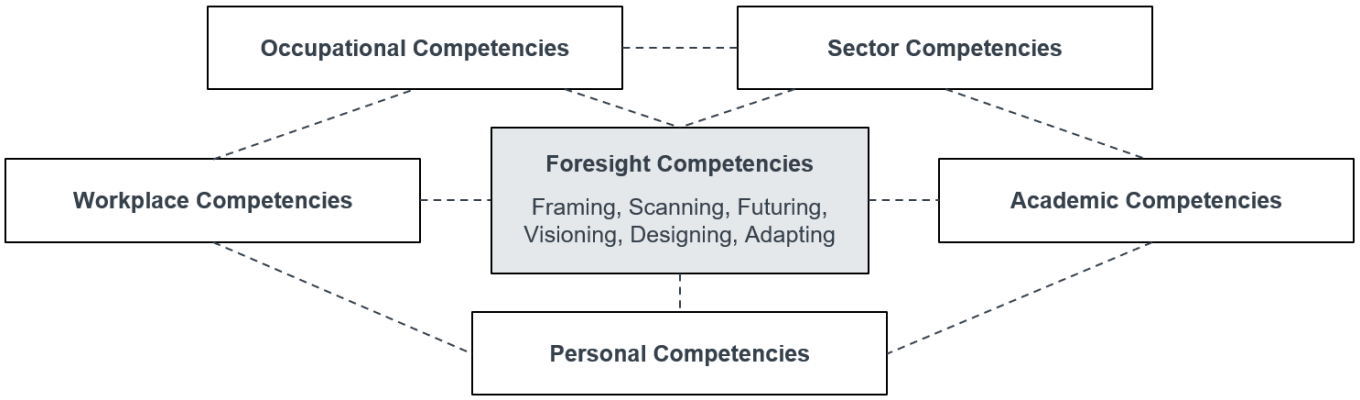


Figure 4: Foresight Competency Model by Hines et al. (2017) (Own Illustration).

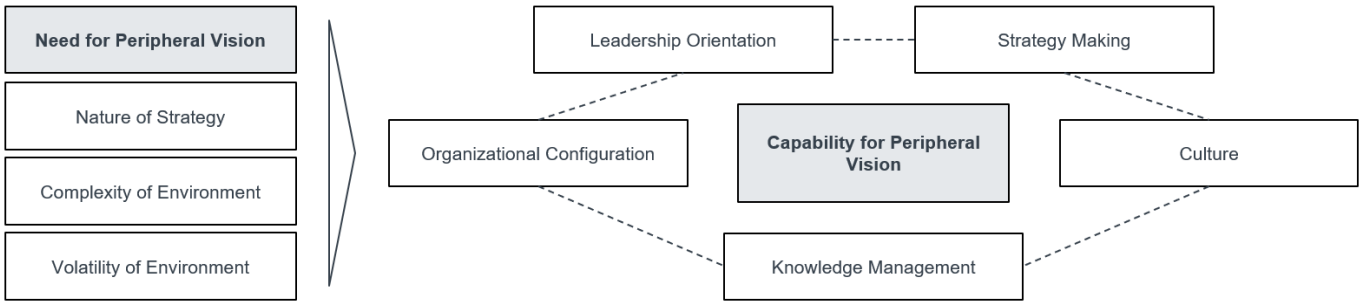


Figure 5: Peripheral Vision Capabilities by Day and Schoemaker (2005) (Own Illustration).

fore, Day and Schoemaker (2005) regard foresight capability as a mediator but imply that firms should not try to maximize the maturity of the mentioned elements (contextual foresight capabilities) but rather adapt it to its specific context in order to attain positive outcomes.

Second, Grim (2009) developed the Foresight Maturity Model (FMM) which defines best practices in foresight activities to assess organizational foresight capability. Those best practices are defined within six different "disciplines" (necessary capabilities) in the strategic foresight process which are leadership, framing, scanning, forecasting, visioning, and planning (Grim, 2009). Those disciplines are inspired by the early work of Hines et al. (2017) "Thinking about the Future" who later co-developed the previously introduced Foresight Competency Model (FCM). However, Grim (2009) covers those disciplines on an organizational level while Hines et al. (2017) described them as individual ("core") capabilities. In the FMM, Grim (2009) provides up to five best practices for each discipline and a maturity index with five levels. Those levels range from "ad hoc" to "world-class" maturity and aim to measure the performance and capability in each practice (Grim, 2009). This differentiation can be seen in Figure 6 below.

Foresight capability is then assessed by adding up the lowest score of each discipline to a total numerical score (Grim, 2009). This suggests that each discipline is only as mature as its weakest practice (Grim, 2009). Therefore, the Foresight Maturity Model defines foresight capability as a state that is achieved by developing high maturity among all relevant practices and disciplines (Grim, 2009). Overall, the model developed by Grim (2009) identifies necessary processual capabilities (disciplines) and describes their components in more detail (best practices). In contrast to Day and Schoemaker (and more recent models), the FMM suggests that maximized proficiency (instead of relative) leads to positive performance implications (Grim, 2009). However, apart from this significant difference, the identified necessary capabilities (disciplines) are very similar to the suggestions of Day and Schoemaker (and more recent models).

Third, Rohrbeck (2010) created the Maturity Model of Corporate Foresight to further advance the assessment and development of organizational foresight capability. This model builds on insights from the existing models and adds complementing criteria identified in practice (Rohrbeck, 2010). As illustrated in Figure 7, the framework consists of three parts which are context, capabilities, and impact of foresight activities. "Context" is based on the idea of Day and Schoemaker (2005) that an organization's foresight requirements depend on its surroundings (Rohrbeck, 2010). However, further components like the size of the company, the corporate culture, and competitive dynamics are added (Rohrbeck, 2010). "Capabilities" are necessary abilities for successful foresight activities and are assessed on their maturity for dealing with discontinuous change (Rohrbeck, 2010). The different capabilities in this model are (1) information usage, (2) method sophistication, (3) people and networks,

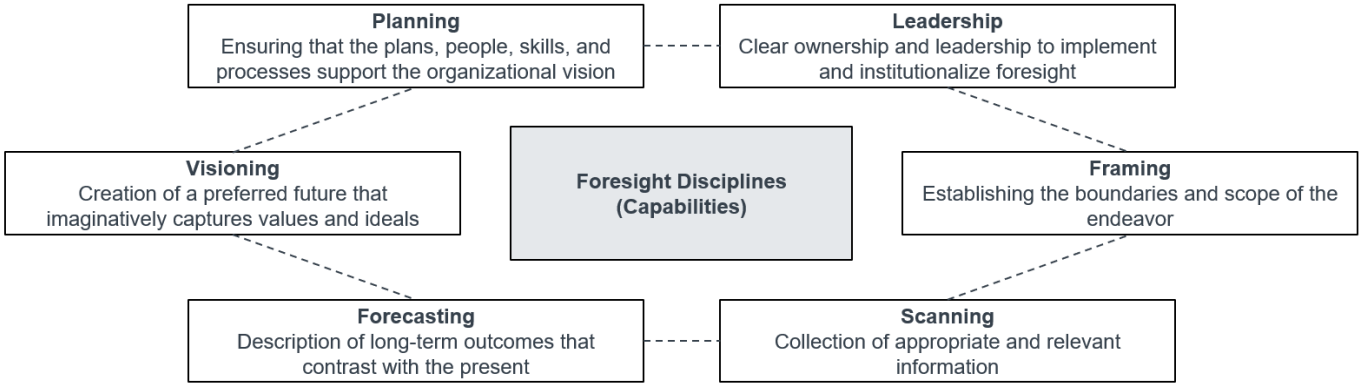


Figure 6: Maturity Index by Grim (2009) (Own Illustration).

(4) organization, and (5) culture (Rohrbeck, 2010). The maturity of each of those rather contextual capabilities is measured and summed up to draw conclusions regarding the foresight capability of a firm (Rohrbeck, 2010). The insights of this measurement are then used to assess foresight activities' "impact" and "value contribution" on firm performance to conclude whether foresight capability has a facilitating or inhibiting effect (Rohrbeck, 2012; Rohrbeck & Schwarz, 2013).

Overall, this model advances the existing frameworks by directly connecting foresight capability to foresight's success and impact on firm performance (Rohrbeck, 2012). Therefore, it does not only implicitly suggest foresight capability as a mediator for firm performance but conceptually and practically showcases the activity-outcome relationship. While this maturity model also emphasizes the importance of foresight's context, the proposed necessary capabilities differ from earlier models and focus on contextual rather than processual capabilities.

Most recently, Rohrbeck and Kum (2018) developed the Future Preparedness Model (FP) to further advance the conceptual models on organizational foresight capability. In this model, they tried to improve the measurability of existing elements and increase the link to firm performance (Rohrbeck & Kum, 2018). The model assesses two components which are the maturity and the need for foresight. The attained relative levels of those elements define the "future preparedness" of an organization (Rohrbeck & Kum, 2018). The maturity element builds directly on the Maturity Model of Rohrbeck (2010) but separates "people and networks" into two components and adds a supporting process layer (Rohrbeck & Kum, 2018). The processes in this layer are perceiving, prospecting, and probing and they group the practices (formerly capabilities) into three categories (Rohrbeck & Kum, 2018) as displayed in Figure 7.

On the other hand, the "foresight need" is assessed with an approach similar to Day and Schoemaker (2005) but normalized to a four-level scale (Rohrbeck & Kum, 2018). In combination, the optimum future preparedness is achieved when the needed foresight level is equal to the attained maturity level. Deviations from this optimum can occur with a maturity that is either lower or higher than needed (Rohrbeck & Kum, 2018). Like Peripheral Vision Capabilities by Day and Schoemaker (2005), the model emphasizes that both, a lack of foresight and too much foresight, can harm firm performance (Rohrbeck & Kum, 2018). Once again, foresight capability is, therefore, not determined only by the maturity of foresight activities but by the match of need and maturity. Overall, the Future Preparedness Model provides adapted capabilities compared to the Maturity Model by Rohrbeck (2010) and integrates processual as well as contextual elements. Also, it directly links foresight capability to firm performance and empirically assesses its impact in a longitudinal study (Rohrbeck & Kum, 2018). This study identifies foresight capability as a mediator for foresight's success and shows that attained capability can lead to higher profitability and market share growth (Rohrbeck & Kum, 2018).

This concludes the existing findings on foresight capability. Research suggests that certain processual and contextual capabilities can facilitate successful foresight and lead to positive outcomes (e.g., Rohrbeck, 2012; Rohrbeck & Kum, 2018; Yoon et al., 2018). Foresight capability is identified as a mediator for foresight's impact on firm performance. However, it remains unclear how foresight's foundations and micro-foundations affect organizational foresight capability and how individual capabilities translate to an organizational level. Further, the exact mediating impact on firm performance is unclear because of relevant moderators and limited empirical studies. To increase transparency, those moderators are discussed in the next chapter.

4.2.4. Moderators for Strategic Foresight's Success

While foresight capability influences the outcomes of foresight practices, some moderators also positively or negatively affect this relationship. According to Sarpong, Maclean, and Alexander (2013), they can either be facilitators or inhibitors for positive outcomes from strategic foresight and are, therefore, important to consider in practice. Those moderators are broadly separated into structural, cultural, and technological moderators in this paper and are displayed in detail below.

First, structure-related moderators capture all elements

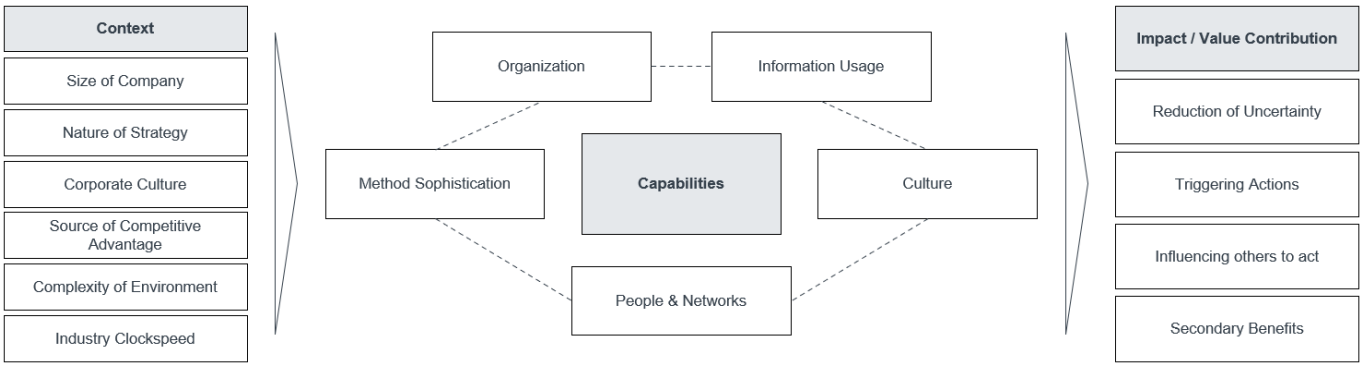


Figure 7: Maturity Model of Corporate Foresight of Rohrbeck (2010) (Own Illustration).

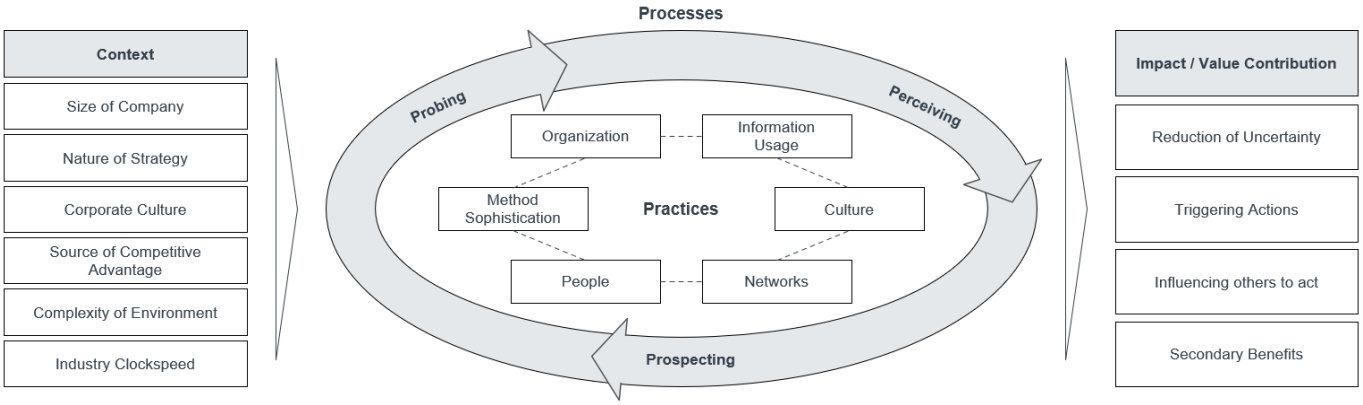


Figure 8: Future Preparedness Model by Rohrbeck and Kum (2018) (Own Illustration).

related to organizational design or process structures. Existing research displays that such formal structures influence the outcomes of foresight practices (e.g., Daheim & Uerz, 2008; Peter & Jarratt, 2015). Regarding organizational design, findings suggest that flat hierarchies facilitate positive outcomes while complex, hierarchical designs can impede foresight's success (Costanzo, 2004). Further, research emphasizes that foresight's impact is improved by institutionalizing foresight activities (e.g., in separate organizational units) since this allows unbiased processes with little dependence on individuals (Milshina & Vishnevskiy, 2018; Rohrbeck & Kum, 2018; Wiener et al., 2020). Meanwhile, research also suggests that some flexibility is necessary since strictly institutionalized foresight practices can lead to organizational blindness that results in foresight failures and further disadvantages (Costanzo, 2004; Ruff, 2015). Regarding process structures, research suggests that the linkage of foresight activities and strategic processes crucially affects foresight outcomes (Farrington et al., 2012; Rohrbeck & Kum, 2018). In particular, missing structure and integration negatively influences achieved results and can cause uncertainty regarding foresight legitimation (Daheim & Uerz, 2008; Milshina & Vishnevskiy, 2018). This can lead to dissatisfaction among involved individuals and cause organizational inertia (Daheim & Uerz, 2008; Haarhaus & Liening, 2020; Vecchiato, 2020). In contrast, research shows that if foresight is embedded in strategic decision-making and backed by the management, final outcomes are improved (e.g., Battistella, 2014; Rohrbeck & Schwarz, 2013; Tapinos, 2013).

Second, culture-related moderators comprise the openness of organizational culture, shared values, and managerial mindsets as elements that affect foresight's outcomes. Regarding cultural openness, research suggests that open-minded cultures facilitate successful foresight and positive implications while a restrictive culture can have a negative effect (Daheim & Uerz, 2008; Major & Cordey-Hayes, 2000; Ruff, 2006). This is in line with the finding that more communication and collaborative foresight positively influence the achieved outcomes (Haarhaus & Liening, 2020; Sarpong & Maclean, 2016; Savioz & Blum, 2002; Wiener et al., 2020). Further, research on shared values in strategic foresight processes emphasizes that they moderate the outcomes of foresight activities by positively or negatively affecting them. Several publications show that organizations require a shared set of values to develop a basis for successfully approaching the future together (Gattringer & Wiener, 2020; Sarpong, Maclean, & Davies, 2013). Here, different views and values are still possible as long as organizations have a shared overarching mindset (Boe-Lillegraven & Monterde, 2015). If this is not the case, different values can inhibit successful foresight and be a disadvantage (Sarpong, Maclean, & Davies, 2013). Lastly, research emphasizes the importance

of managerial mindsets for strategic foresight outcomes (Li & Sullivan, 2022). Findings suggest that a positive managerial attitude toward foresight, change, and the future enables positive outcomes from foresight activities (e.g., Haarhaus & Liening, 2020; Klos & Spieth, 2021; Peter & Jarratt, 2015; Rohrbeck & Schwarz, 2013). In contrast, negative mindsets and past-focused management styles negatively impact foresight's success (Haarhaus & Liening, 2020; Major & Cordey-Hayes, 2000). Research points out that the mindsets of middle managers are particularly important since they actively implement the strategies developed through foresight and, therefore, heavily affect the impact on organizations and their performance (Darkow, 2015; Sarpong & Hartman, 2018).

Third, technology-related moderators play an increasingly important role for strategic foresight activities and their outcomes (Von der Gracht et al., 2015). Research shows that technology can facilitate positive results of foresight practices by making them more transparent and collaborative (Rohrbeck, Battistella, & Huizingh, 2015; Von der Gracht et al., 2015). More precisely, technology can open foresight activities up to more participants, ease communication between them, and facilitate a supportive culture for successful foresight (Rohrbeck, Battistella, & Huizingh, 2015). This suggests that the moderating effect of technology is rather indirect and influences the activity-outcome relationship of foresight by affecting structure and culture (Marinković et al., 2022). In addition to this indirect effect, technology also directly moderates the outcomes of foresight by improving the processes of foresight activities: research emphasizes that modern technology can improve the quality and efficiency of foresight activities, especially when they are data-driven (Yoon et al., 2019). Disparately, technology is not only a moderator for foresight processes but can also be the initial motivation for such practices or a (positive) outcome in itself (Marinković et al., 2022). Broadly, this can be seen in the organizing framework introduced in chapter 4.1.3. Additionally, details on technological outcomes of foresight are presented in chapter 4.3.2.

4.3. Strategic Foresight's Impact on Firm Performance

The range of strategic foresight's implications for firm performance is broad and includes external as well as internal outcomes. The differentiation of those outcomes made in the organizing framework (chapter 4.1.3) is once again shown in the following table (Table 6). The findings regarding those different categories will be displayed in detail in the following sections. This display aims to provide the basis to answer the third and final research question of this paper Q3 ("How does successful strategic foresight reflect itself in overall firm performance?").

4.3.1. Strategic Impact

One research stream in the strategic foresight field shows that successful foresight practices can improve decision-making, strategic planning as well as overall strategic agility and flexibility (e.g., Rohrbeck, 2012; Ruff, 2015; Vecchiato, 2015). The detailed findings on such strategy-related, qualitative outcomes are presented and discussed in the paragraphs below.

Firstly, several publications emphasize the idea that strategic foresight positively impacts decision-making (e.g., Battistella & De Toni, 2011; Heger & Rohrbeck, 2012; Ringland, 2010). As explained in the introduction, understanding the implications of environmental changes to adapt decision-making is one of the theoretical motivations for organizations to employ foresight practices (Calof et al., 2018; Yoon et al., 2018). Research demonstrates that this can be achieved in practice and attribute this to strategic foresight's outside-in perspective which allows firms to identify trends and changes (before the competition) (Peter & Jarratt, 2015; Rohrbeck & Schwarz, 2013). As a result, decision failures are reduced significantly in firms with mature strategic foresight practices (Chermack, 2004). While some scholars argue that this impact is too generic and hard to quantify (Iden et al., 2017), research does support foresight's positive impact on decision-making through several qualitative studies (e.g., Gershman et al., 2016; Heger & Rohrbeck, 2012; Rohrbeck, 2012).

Secondly, findings showcase that foresight practices can improve strategic planning (e.g., Ringland, 2010; Von der Gracht & Stillings, 2013; Yoon et al., 2019). This is attributed to the fact that strategic foresight makes the future more tangible through developing and illustrating potential scenarios (Lehr, Lorenz, Willert, & Rohrbeck, 2017; Weber et al., 2015). In particular, distinct scenarios and roadmaps developed in foresight activities are believed to have a positive impact on strategic planning (Milshina & Vishnevskiy, 2018; Yoon et al., 2019). As indicated through the organizing framework (chapter 4.1.3), technological aids and process tools are moderators that can facilitate foresight's success and improve strategic planning as well as strategy development (Farrukh & Holgado, 2020; Tapinos, 2013). While researchers agree on a positive outcome here, exact effects are only observable in the long term (Iden et al., 2017; Rohrbeck, 2012).

Thirdly, the findings of scholars show that foresight can positively impact the strategic flexibility and agility of organizations. This is again connected to the outside-in perspective of foresight that allows firms to better anticipate and respond to external changes (Battistella, 2014; Peter & Jarratt, 2015). Here, several researchers point out that strategic foresight can lead to significant flexibility/agility improvements (Gershman et al., 2016; Haarhaus & Liening, 2020; Vecchiato, 2015), due to increased organizational reactiveness that results from a better environmental understanding (Battistella, 2014; Battistella & De Toni, 2011; Rohrbeck, 2012; Vecchiato & Roveda, 2010b). Studies show that this improved ability to (proactively) respond to change is connected to enhanced organizational learning that results from foresight activities (Rohrbeck, 2012). This concludes the findings on the three main strategic impacts that successful strategic foresight can have on (internal) firm performance.

Table 6: Foresight's Impact on Firm Performance Research Structure.

Research Theme	Topic	Influential Studies
Strategic Impact	I. Strategic Decision-Making	Yoon et al., 2019; Schwarz et al., 2019; Milshina & Vishnevskiy, 2018; Gershman et al., 2016; Von der Gracht & Stillings, 2013; Rohrbeck, 2012; Ringland, 2010
	II. Strategic Planning	Yoon et al., 2019; Gershman et al., 2016; Peter & Jarratt, 2015; Von der Gracht & Stillings, 2013; Rohrbeck, 2012; Ringland, 2010
	III. Strategic Agility / Flexibility	Haarhaus & Liening, 2020; Wiener et al., 2020; Gershman et al., 2016; Peter & Jarratt, 2015; Vecchiato, 2012
Organizational Impact	I. Organizational Learning, Consensus and Communication	Wiener et al., 2020; Burt & Nair, 2020; Schweitzer et al., 2019; Bootz et al., 2019; Ruff, 2015; Paliokaitė & Pačėsa, 2015; Boe-Lillegraven & Monterde, 2015; Van der Duin et al., 2014; Rohrbeck, 2012; Ringland, 2010; Costanzo, 2004
	II. Technology Innovation	Wiener et al., 2020; Schwarz et al., 2019; Ho & O'Sullivan, 2018; Gershman et al., 2016; Scheiner, Baccarella, Bessant, & Voigt, 2015; Rohrbeck, Battistella, & Huizingh, 2015; Paliokaitė & Pačėsa, 2015; Battistella, 2014; Von der Gracht & Stillings, 2013; Rohrbeck & Gemünden, 2011
	III. Portfolio Innovation	Schweitzer et al., 2019; Ruff, 2015; Vecchiato, 2012; Battistella & De Toni, 2011
Performance Impact	I. Competitiveness	Arokodare & Asikhia, 2021; Eskandari et al., 2020; Ho & O'Sullivan, 2018; Nkuda, 2017; Vecchiato, 2015; Reid & Zyglidopoulos, 2004
	II. Profitability	Arokodare & Asikhia, 2021; Rohrbeck & Kum, 2018; Boe-Lillegraven & Monterde, 2015; Rohrbeck & Schwarz, 2013; Vecchiato, 2012; Rohrbeck, 2012

4.3.2. Organizational Impact

The second cluster of research regarding foresight's impact on firm performance focuses on the organization- and innovation-related outcomes of foresight practices. Research suggests that it facilitates organizational change and influences how individuals work (together) (Rohrbeck, 2012; Paliokaitė & Pačėsa, 2015; Van der Duin et al., 2014), while those changes and the outside-in orientation of foresight can improve firms' innovation capabilities (Adegbile et al., 2017; Paliokaitė & Pačėsa, 2015). The findings are differentiated into organizational change, technological innovation, and portfolio innovation will be displayed below.

While foresight's organizational impact is often regarded as a side effect, the perceived importance of such outcomes is growing rapidly (Marinković et al., 2022). Research, for example, suggests that foresight's sensitization to environmental changes induces increased reflection and organizational renewal (Burt & Nair, 2020; Ruff, 2015; Wiener et al., 2020). Also, research shows that the collaborative tools used for foresight can improve communication and consensus within organizations (Bootz et al., 2019; Ramírez et al., 2013; Rohrbeck & Schwarz, 2013). On one hand, this is the case because foresight requires communication within an organization and stimulates discussions (Rohrbeck, 2012). On the other hand, foresight tools often integrate multiple stakeholders and open the discussion for members from different hierarchy levels which increases collaboration and consensus according to scholars (Boe-Lillegraven & Monterde, 2015; De Smedt et al., 2013; Ho & O'Sullivan, 2017). Overall, research shows that foresight can improve organizational change, communication, and consensus which positively impacts organizational efficiency and effectiveness (Ilmola & Kuusi, 2006; Wiener et al., 2020).

Generally, studies show that strategic foresight has a positive impact on the ambidexterity of firms (Paliokaitė & Pačėsa, 2015; Sarpong, Maclean, & Davies, 2013). Focused on technological innovation, one research stream attests that early assessment of new technologies, weak signal identification, and the improved ability to create radical innovation are positive impacts of foresight practices (Gershman et al., 2016; Rohrbeck, Battistella, & Huizingh, 2015; Scheiner et al., 2015; Wiener et al., 2020). A different research stream implies that firms with foresight capability are better at strategically integrating technological innovations in their businesses due to better process understanding (Battistella, 2014; Rohrbeck, Battistella, & Huizingh, 2015; Paliokaitė & Pačėsa, 2015; Schwarz et al., 2019). Therefore, foresight can have a positive impact on the technological innovation capabilities of a firm due to an improved understanding of its processes and surroundings.

Lastly, research indicates that strategic foresight can have a positive effect on portfolio innovation which includes products, business models, and markets (Von Der Gracht, Vennemann, & Darkow, 2010). Here, research points out that successful foresight can improve organizations' capability to develop new products and fulfill customer needs through envisioning scenarios with future preferences (Açikgöz, Günsel, Kuzey, & Zaim, 2016; Ho & O'Sullivan, 2018; Schwarz et al., 2019; Un & Price, 2007; Wright et al., 2008). Further, scholars show that firms with foresight capability develop a better

market understanding and are more successful in identifying new business fields due to environmental scanning activities (Boe-Lillegraven & Monterde, 2015; Fritzsche, 2018; Rohrbeck & Gemünden, 2011). Also, research suggests that incorporating foresight in the creation and evaluation of business models can improve the result (Farrington et al., 2012; Højland & Rohrbeck, 2018; Van der Duin et al., 2014). In sum, strategic foresight can, therefore, have a significantly positive impact on an organization's portfolio innovation.

4.3.3. Performance Impact

Due to strategic foresight's qualitative nature and long-term orientation, most of the existing research focused on internal, qualitative outcomes (Boe-Lillegraven & Monterde, 2015; Vecchiato, 2012). However, recent studies attempted to show the external performance impact of foresight activities by assessing their effect on the competitiveness and profitability of firms (e.g., Rohrbeck & Kum, 2018). The findings in those two categories are displayed below.

In practice, it is very difficult to isolate the quantitative impact of foresight (Rohrbeck, 2012; Rohrbeck & Schwarz, 2013). Therefore, many studies on foresight's impact on external firm performance rather focused on non-financial competitiveness (Milshina & Vishnevskiy, 2018; Wiener et al., 2020). According to those studies, firms that apply strategic foresight perform better compared to their competition which is attributed to the strategic and organizational benefits of foresight activities that were introduced above (Marinković et al., 2022). One study, for example, suggests that strategic foresight enhances competitiveness through improved organizational efficiency and optimized offerings (Eskandari et al., 2020; Ho & O'Sullivan, 2018) while other scholars point out that the strategic agility that results from foresight practices can lead to sustainable competitive advantages (Arokodare & Asikhia, 2021; Nkuda, 2017; Vecchiato, 2015). Inversely, research suggests that a lack of foresight can lead to negative competitive outcomes (Reid & Zyglidopoulos, 2004). In conclusion, research, therefore, indicates that successful strategic foresight can improve the competitive dynamics of an organization.

Recently, a new research stream further investigated the quantitative, financial impact of strategic foresight on organizational profitability (Rohrbeck, 2012; Rohrbeck & Kum, 2018). This research indicates that strategic foresight activities can be a good investment and increase firm profitability over time (Rohrbeck, 2012). However, since outcomes are mostly observable in the long term, investments in foresight are often neglected in favor of other ventures (Iden et al., 2017; Rohrbeck, 2012). To capture those long-term effects, a longitudinal study was conducted which attests that firms with foresight capability show up to "33% higher profitability and a 200% higher market capitalization growth" than the average of compared firms (Rohrbeck & Kum, 2018). Other studies also suggest that foresight activities can lead to superior firm performance and market share growth (Arokodare & Asikhia, 2021). In these studies, researchers generally see foresight's impact on profitability as an indirect result of the mentioned qualitative outcomes (Marinković et al., 2022). Consequently, there still seems to be a lack of evidence for the direct effects of foresight on profitability and competitiveness which needs to be addressed in future research. Still, first studies already exhibit empirically that successful foresight can improve quantitative, financial performance (Rohrbeck & Kum, 2018).

This concludes the potential impacts strategic foresight (capability) can have on firm performance. For many of those outcomes, measurement is rather difficult and the root cause for the (positive) effect is uncertain (Iden et al., 2017; Rohrbeck, 2012). Still, research clearly suggests that foresight capability can be a facilitator for successful foresight and positive implications for firm performance (e.g., Gershman et al., 2016; Heger & Rohrbeck, 2012). Nevertheless, further (empirical) research is necessary to quantify exact outcomes and make foresight's impact on firm performance more tangible. This can help practitioners in directing their efforts to certain topics in foresight practices and advance academia through more empirical evidence.

5. Discussion

The preceding chapters descriptively displayed the current state and knowledge on different topics in strategic foresight research. This chapter builds on this display and discusses the provided information to answer the paper's research questions (Q1-Q3). In doing so, this chapter aims to point out theoretical and practical implications as well as the limitations of the findings. Overall, the systematic literature review in this paper offers an updated, more extensive picture of existing research, structures the field through the organizing framework, and provides a state-of-the-art foundation for future research on foresight capability and foresight's impact on firm performance. The following paragraphs explain those contributions in detail.

5.1. Theoretical and Practical Implications

First, the quantitative review in this paper organized existing foresight research and showcased trends. Regarding the question of what the current state of strategic foresight research is (Q1), the following two insights were gathered: (1) Research on strategic foresight is becoming more relevant in the scientific discourse (increased yearly publications in higher-ranked journals) but must still develop its foundations. Many publications don't have a clear theoretical foundation (n=88) and are untransparent regarding the analyzed epistemological dimension (individual- or organization-level). (2) The organizing model developed in this paper points out that existing research mostly focused on antecedents and foundations of foresight (n=191) while findings on capability, moderators, and firm performance are still scarce (n=75) despite their practical importance. Therefore, this paper identifies those themes as relevant trajectories for future research and suggests this as an opportunity to transition foresight into management journals.

An important contribution of this quantitative part is the comprehensive overview of existing research that was created based on a sample of 243 journal-published articles. While prior research included some literature reviews (e.g., Iden et al., 2017; Singh et al., 2020), the field lacked transparency and research was never fully consolidated. Therefore, this paper extends prior work by providing the most comprehensive and most recent picture of the field to date which supports the effort of developing one unified understanding of strategic foresight. A second contribution is the developed organizing framework of strategic foresight research that displays relevant research themes and their conceptual connections. Prior research was unstructured and had not identified a clear pattern of studies (Hines, 2020; Iden et al., 2017; Rohrbeck, Battistella, & Huizingh, 2015; Snyder, 2019). Thus, the organizing framework advances the field by uniquely consolidating the most relevant research themes, displaying their connections in a clearly structured manner, and pointing out trajectories for future research. Further, this display uncovers the importance of mediators (foresight capability) and moderators for foresight's impact on firm performance which adds an important research frontier. Overall, the quantitative review creates a better understanding of the foresight field and enables more targeted future research.

Second, the qualitative review on strategic foresight capability displayed and contrasted existing conceptual models on foresight capability to understand its components and outcomes. Regarding the question of which capabilities are necessary for successful foresight (Q2), the following three findings were made: (1) Conceptual models to date do not have a unified understanding of foresight capability and differentiate between organizational or individual capabilities. While some studies suggest that individual capability can potentially result in an isomorphic organizational capability, this hypothesis must be viewed skeptically since it lacks empirical evidence. (2) Different models suggest different necessary capabilities for successful strategic foresight. Those capabilities can be grouped into processual (e.g., framing, scanning, forecasting, etc.) and contextual capabilities (organizational configuration, culture, leadership, etc.). Developing a combination seems most fitting to ensure successful foresight in practice. (3) Several models imply that the maturity of those capabilities should not be maximized but rather adapted to an organization's environment and context. Contingently, highly developed capabilities can have a negative impact on firm performance if the environmental requirements for foresight are low. Matching maturity and need is expected to result in the best outcomes.

One contribution of this qualitative review is the raised awareness regarding different epistemological dimensions of foresight capability. Prior studies often neglected this discrepancy and built on studies that had different units of analysis (e.g., Hines et al., 2017; Grim, 2009). Therefore, this paper advances foresight studies by enabling a more conscious and differentiated handling of preceding studies which can improve the quality of overall foresight research. Also, this paper contributes to foresight studies by extensively consolidating the current knowledge on foresight capability. Prior to this, different capability models were published in isolation and not compared or contrasted with each other (e.g., Day & Schoemaker, 2005; Grim, 2009; Rohrbeck, 2010). This paper, therefore, extends prior research by uncovering similarities and differences between existing models and by categorizing necessary capabilities into processual and contextual elements. Overall, this is the first literature review that extensively showcases findings on distinct capabilities that compose foresight capability which lays the foundation for a productive discourse about foresight capability. However, the suggested capabilities and their potential impact on firm performance were derived from conceptual studies and have often not yet been empirically tested (in isolation). Therefore, while those findings already offer theoretical and practical implications, they must be verified further in future research.

Third, the qualitative review on foresight's impact on firm performance provides a comprehensive list of potential outcomes. Regarding the question of how successful foresight reflects itself in firm performance (Q3), the following two insights were gathered: (1) Successful foresight activities can have a significantly positive impact on internal and external firm performance. This positive impact manifests itself in strategic, organizational, and performance outcomes that can allow firms to gain a competitive advantage and achieve superior performance in volatile business environments. (2) Most existing studies concerning foresight's outcomes build on qualitative research designs and suggest qualitative, long-term results that are difficult to measure. Only very few articles in the paper's sample (n=4) quantify foresight's impact on firm performance and empirically assess it over a longer period. Thus, it is still difficult to determine and quantify the exact, isolated impact of foresight (capability) on firm performance.

The main contribution of this part is the extensive overview of potential qualitative and quantitative impacts of foresight on firm performance. Prior research mostly dealt with qualitative, internal outcomes and lacked a consolidation of all potential impacts (e.g., Gershman et al., 2016; Peter & Jarratt, 2015; Paliokaitė & Pačėsa, 2015; Yoon et al., 2019). Therefore, this paper extends existing research by providing a comprehensive overview that includes internal, qualitative as well as external, quantitative outcomes from foresight (capability). This enables scholars to empirically test and quantify the exact activity-outcome relationship of foresight. Exploring those topics further can advance the field by making the rather qualitative and opaque topics of foresight capability and foresight outcomes more measurable and tangible.

5.2. Limitations and Shortcomings

While the research approach and methodology for this paper were selected carefully, there are limitations that need to be considered to fully understand the paper's implications. The two main limitations are discussed in the following paragraphs to create more transparency.

First, there are limitations regarding the sample of articles used in this paper. The sample included only journal-published articles (with two exceptions) and omitted other sources like books or conference papers. Therefore, the findings presented in those other research outlets were not considered in this paper's analyses which limits the exhaustiveness of its findings. Further, the sample excluded papers published prior to the year 2000 which also limits the finding's exhaustiveness since there might be relevant articles dated earlier. Also, it cannot be guaranteed that all relevant publications were included due to the selective usage of AI-search engines and research databases as well as broadly specified search keywords. Adaptions to those aspects could potentially change the final sample and result in more refined outcomes. However, the most relevant contributions are expected to be included either way due to the reproducing nature of research which suggests that the impact of those limitations is minor.

Second, a limitation of this paper's practical implications arises from the nature of the employed research approach. More precisely, findings on foresight capability mostly build on conceptual articles (e.g., theoretical frameworks, models, etc.) that often do not offer (extensive) empirical evidence. For example, while there are studies that connect foresight capability to firm performance and empirically observe certain outcomes, it cannot fully be inferred how distinct capabilities (and overall foresight capability) influence those outcomes in isolation. Therefore, more empirical research must be done to determine which capabilities are necessary for successful foresight. Similarly, the findings regarding foresight's impact on firm performance are mostly based on qualitative studies that do not allow conclusions on the exact root of outcomes. Since the literature review in this paper only compiled and discussed this information, it does not resolve this problem by empirically verifying the findings. Therefore, the findings do not offer proven implications for practice but should rather be seen as well-founded hypotheses that require verification. Important research trajectories, therefore, are to measure and quantify the outcomes of foresight capability on firm performance as well as the influence of moderators on this relationship. This paper's findings can be used as the foundation for that.

5.3. Evaluation of Methodology and AI Search Engines

The paper used the AI-based search engines "Iris.ai" and "Semantic Scholar" as the primary sources for data gathering and cross-checked the identified articles with the traditional databases Elsevier, JSTOR, and SAGE Journals. This methodology was not only employed to utilize the potential upsides of modern AI engines but also to assess and evaluate the current maturity of two very promising examples. The second objective was pursued by comparing the articles identified by the AI engines to the mentioned databases which are considered very exhaustive (Iden et al., 2017; Marinković et al., 2022). This comparison was done both for the full sample and for the 20 most frequently cited articles according to Dimensions AI (App. C).

Overall, this comparison draws the conclusion that AI engines can be a good addition to existing tools but lack comprehensiveness when used in isolation. While Iris.ai develops visual representations of existing literature that can be very helpful at the start of a research project (see Appendix C), the engine only identified 98 out of the 243 articles relevant for the final sample (40%) including 18 out of the 20 most frequently cited articles (90%). Therefore, this engine alone does not yet seem to be mature enough to enable a complete, extensive research project. Semantic Scholar, on the other hand, covered 159 articles of the full sample (65%) and 19 of the most frequently cited articles (95%) which suggests significantly higher comprehensiveness. Nevertheless, this engine's database does still omit around a third of relevant articles and did not display clear advantages compared to the traditional databases used for cross-checking.

Therefore, the maturity of AI-based search engines as a methodological innovation for literature reviews is currently deemed too low for independent usage. While they do provide some of the advertised benefits, they are still far from exhaustive which can significantly decrease the output quality. If a literature review aims to provide a full, comprehensive overview of a research field, this would be a particularly severe problem. Building on the intensive usage of those engines for this paper, the two core recommendations are (1) to use them only in combination with other, more established engines and (2) to utilize them for specific objectives rather than full studies. For example, the visual "explore maps" created by Iris.ai can be very helpful at the beginning of a study but are insufficient for an extensive data gathering process.

6. Conclusion

The systematic literature review on strategic foresight conducted in this thesis aimed to structure the developing field and build a theoretical basis for future research and practice. This was done through a quantitative display of descriptive trends and a comprehensive organizing framework of existing research as well as qualitative, narrative syntheses of the current knowledge on strategic foresight capability and its impact on firm performance. By answering the three research questions of this paper, it contributes to the effort of advancing the foresight field by enabling more targeted future research and accelerating theoretical progress.

The review covered 243 journal-published articles on strategic foresight between 2000 and August 2022 which is the most extensive and most recent display of research to date. By analyzing those articles, the paper identifies that the field still lacks clear theoretical foundations, mixes different epistemological dimensions (individual and organizational foresight), and builds mostly on explorative case studies or conceptual frameworks that do not work toward a clear, shared objective. Themes in foresight research were identified and used to develop an organizing framework that consists of antecedents, capability, organizational foundations, individual micro-foundations, moderators, and outcomes of foresight. The model points out the connections between the

themes which can be utilized to structure scholars' efforts to collaboratively advance the field. Through this, a unified scientific discourse on strategic foresight can be started which increases the maturity of the field and leads to tenable practical implications.

Also, the review compiles and contrasts existing knowledge on necessary capabilities to ensure successful foresight activities that can grant firms a competitive advantage. Here, the paper finds that distinct processual (e.g., framing, scanning, forecasting, etc.) and contextual (e.g., organizational configuration, culture, leadership, etc.) capabilities should be developed in accordance with an organization's faced environment to ensure success. Since those suggested capabilities build on the assumptions of conceptual frameworks, the paper proposes that those assumptions should be tested empirically through future research. This is also the case with existing research on foresight's impact on firm performance since those findings are often (arbitrarily) derived and aggregated from case studies but not verified and discussed on an overarching scale. Nevertheless, the paper clearly identifies potential positive impacts of strategic foresight on firm performance which manifest in strategic (decision-making, planning, and flexibility), organizational (communication, consensus, change, and innovation), and performance outcomes (competitiveness and profitability) in practice. This supports the underlying hypothesis that strategic foresight practices can provide firms with a competitive advantage by reducing the uncertainty that complex, volatile business environments put upon them.

While there are limitations that arise from the methodology of this paper, it does provide a comprehensive picture of existing research with its trends and gaps. This can guide future research and support practitioners in discussing and implementing foresight. In particular, it points out the following two broad trajectories: (1) Empirical testing of the exact (quantitative) outcomes of foresight practices with all its facets. More precisely, it is still unclear which effects result directly from foresight capabilities and which outcomes are rather indirect results from organizational changes that enable foresight. Also, the impact of moderators has not yet been assessed in isolation which reduces the significance of practical implications. (2) Conceptual and empirical research on foresight capability is another frontier. This paper only identified 9 articles in the sample that primarily focused on this concept and, therefore, built on a very limited foundation even though understanding the capabilities necessary to enable success is highly relevant for organizations that engage in strategic foresight activities. Here, more research must be done to build a unified understanding of foresight capability's components and their empirical foundation. Future research should practically identify and validate necessary capabilities and try to quantify effects that result from overarching foresight capability.

Overall, this paper offers the foundation to tackle those research frontiers and provides indicative practical implications. The mentioned influential studies (e.g., in Table 5 and Table 6) can be used by scholars to build on others' contributions and work toward a combined objective. Through this, future research can be more structured and collaboratively advance the still developing strategic foresight field. This will support the effort of integrating strategic foresight research into discussions in respected management journals and improve research and practice.

References

Açikgöz, A., Günsel, A., Kuzey, C., & Zaim, H. (2016). Team foresight in new product development projects. *Group Decision and Negotiation*, *25*(2), 289.

Adegbile, A., Sarpong, D., & Meissner, D. (2017). Strategic foresight for innovation management: A review and research agenda. *International Journal of Innovation and Technology Management*, *14*(04), 1750019.

Ahlqvist, T., & Kohl, J. (2016). Constructing and mobilising futures knowledge in an organisation: foresight as a catalyst of dynamic strategic practice. *Technology Analysis & Strategic Management*, *28*(10), 1138–1151.

Aichouni, M., Touahmia, M., Kolsi, L., Alghamdi, A. S., & Al-Homaid, T. (2021). Foresight readiness assessment for saudi organizations. *Journal of Futures Studies*, *25*(3), 49–64.

Amsteus, M. (2008). Managerial foresight: concept and measurement. *foresight*, *10*(1), 53–66.

Arokodare, M. A., & Asikhia, O. (2021). Strategic agility: Achieving superior organizational performance through strategic fore-sight. *International Journal of Research in Finance and Management*, *4*(1).

Arokodare, M. A., Makinde, G. O., & Fakunmoju, S. K. (2020). Strategic Agility and Competitive Advantage of Oil and Gas Marketing Companies: The Moderating Effect of Information Technology Capability and Strategic Foresight. *International Journal of Advanced Research*, *8*(8), 1393–1406.

Bailey, M. D. (2006). The meanings of magic. *Magic, Ritual, and Witchcraft*, *1*(1), 1–23.

Ballandonne, M. (2020). The history of futures studies: A note on gilfillan's early work. *Technological Forecasting and Social Change*, *157*, 119983.

Battistella, C. (2014). The organisation of corporate foresight: A multiple case study in the telecommunication industry. *Technological Forecasting and Social Change*, *87*, 60–79.

Battistella, C., & De Toni, A. F. (2011). A methodology of technological foresight: A proposal and field study. *Technological Forecasting and Social Change*, *78*(6), 1029–1048.

Bennett, N., & Lemoine, G. J. (2014). What a difference a word makes: Understanding threats to performance in a vuca world. *Business horizons*, *57*(3), 311–317.

Berger, G., de Bourbon Busset, J., & Massé, P. (2007). *De la prospective: Textes fondamentaux de la prospective française, 1955-1966*. Editions L'Harmattan.

Bezold, C. (2010). Lessons from using scenarios for strategic foresight. *Technological forecasting and social change*, *77*(9), 1513–1518.

Boe-Lillegraven, S., & Monterde, S. (2015). Exploring the cognitive value of technology foresight: The case of the cisco technology radar. *Technological Forecasting and Social Change*, *101*, 62–82.

Bootz, J.-P. (2010). Strategic foresight and organizational learning: A survey and critical analysis. *Technological forecasting and social change*, *77*(9), 1588–1594.

Bootz, J.-P., Monti, R., Durance, P., Pacini, V., & Chapuy, P. (2019). The links between french school of foresight and organizational learning: An assessment of developments in the last ten years. *Technological Forecasting and Social Change*, *140*, 92–104.

Burns, A. (2021). Foundation professor of foresight: Richard slaughter's scholarly book contributions to futures studies and strategic foresight. *Futures*, *132*, 102787.

Burt, G., & Nair, A. K. (2020). Rigidities of imagination in scenario planning: Strategic foresight through 'unlearning'. *Technological Forecasting and Social Change*, *153*, 119927.

Calof, J., Arcos, R., & Sewdass, N. (2018). Competitive intelligence practices of european firms. *Technology Analysis & Strategic Management*, *30*(6), 658–671.

Chermack, T. J. (2004). Improving decision-making with scenario planning. *Futures*, *36*(3), 295–309.

Costanzo, L. A. (2004). Strategic foresight in a high-speed environment. *Futures*, *36*(2), 219–235.

Daheim, C., & Uerz, G. (2008). Corporate foresight in europe: from trend based logics to open foresight. *Technology Analysis & Strategic Management*, *20*(3), 321–336.

Darkow, I.-L. (2015). The involvement of middle management in strategy development—development and implementation of a foresight-based approach. *Technological Forecasting and Social Change*, *101*, 10–24.

Day, G. S., & Schoemaker, P. J. (2005). Scanning the periphery. *Harvard business review*, *83*(11), 135.

De Smedt, P., Borch, K., & Fuller, T. (2013). Future scenarios to inspire innovation. *Technological forecasting and social change*, *80*(3), 432–443.

Djuricic, K., & Bootz, J.-P. (2019). Effectuation and foresight–an exploratory study of the implicit links between the two concepts. *Technological Forecasting and Social Change*, *140*, 115–128.

Dufva, M., & Ahlqvist, T. (2015). Elements in the construction of future-orientation: A systems view of foresight. *Futures*, *73*, 112–125.

Eskandari, F., Mohammadi, E., & Rahimi, H. (2020). The effect of strategic foresight on competitiveness, with the mediating role of strategic learning (case study: Private banks of kermanshah city). *Technium Soc. Sci. J.*, *11*, 276.

Extance, A. (2018). How ai technology can tame the scientific literature. *Nature*, *561*(7722), 273–275.

Farrington, T., Henson, K., & Crews, C. (2012). Research foresights: The use of strategic foresight methods for ideation and portfolio management. *Research-Technology Management*, *55*(2), 26–33.

Farrukh, C., & Holgado, M. (2020). Integrating sustainable value thinking into technology forecasting: A configurable toolset for early stage technology assessment. *Technological Forecasting and Social Change*, *158*, 120171.

Fritzsche, A. (2018). Corporate foresight in open laboratories–a translational approach. *Technology Analysis & Strategic Management*, *30*(6), 646–657.

Gattringer, R., & Wiener, M. (2020). Key factors in the start-up phase of collaborative foresight. *Technological Forecasting and Social Change*, *153*, 119931.

Gattringer, R., Wiener, M., & Strehl, F. (2017). The challenge of partner selection in collaborative foresight projects. *Technological Forecasting and Social Change*, *120*, 298–310.

Gershman, M., Bredikhin, S., & Vishnevskiy, K. (2016). The role of corporate foresight and technology roadmapping in companies' innovation development: The case of russian state-owned enterprises. *Technological Forecasting and Social Change*, *110*, 187–195.

Ghayoor, H., Rastegari, H., & Hosseini, S. H. (2020). Designing and developing the strategic foresight capabilities of medical universities in the accidents and crises of the country with emphasis on human resource agility. *International Journal of Logistics Systems and Management*, *36*(1), 124–137.

Gibson, E., Dime, T., Garces, E., & Dabich, M. (2018). Bibliometric analysis as a tool for identifying common and emerging methods of technological foresight foresight. *V*, *12*, 6–24.

Godet, M. (2000). The art of scenarios and strategic planning: tools and pitfalls. *Technological forecasting and social change*, *65*(1), 3–22.

Gordon, A. V., Ramic, M., Rohrbeck, R., & Spaniol, M. J. (2020). 50 years of corporate and organizational foresight: Looking back and going forward. *Technological Forecasting and Social Change*, *154*, 119966.

Greenblott, J. M., O'Farrell, T., Olson, R., & Burchard, B. (2019). Strategic foresight in the federal government: a survey of methods, resources, and institutional arrangements. *World futures review*, *11*(3), 245–266.

Grim, T. (2009). Foresight maturity model (fmm): Achieving best practices in the foresight field. *Journal of Futures Studies*, *13*(4), 69–80.

Haarhaus, T., & Liening, A. (2020). Building dynamic capabilities to cope with environmental uncertainty: The role of strategic foresight. *Technological Forecasting and Social Change*, *155*, 120033.

Hamel, S., Ims, R. A., & Yoccoz, N. (2022). Challenges and opportunities when implementing strategic foresight: lessons learned when engaging stakeholders in climate-ecological research. *Climate Research*, *86*, 29–35.

Heger, T., & Boman, M. (2015). Networked foresight—the case of eit ict labs. *Technological Forecasting and Social Change*, *101*, 147–164.

Heger, T., & Rohrbeck, R. (2012). Strategic foresight for collaborative exploration of new business fields. *Technological Forecasting and Social Change*, *79*(5), 819–831.

Hines, A. (2020). When did it start? origin of the foresight field. *World Futures Review*, *12*(1), 4–11.

Hines, A., Gary, J., Daheim, C., & van Der Laan, L. (2017). Building foresight capacity: toward a foresight competency model. *World Futures Review*, *9*(3), 123–141.

Hines, A., & Gold, J. (2015). An organizational futurist role for integrating foresight into corporations. *Technological Forecasting and Social Change*, *101*, 99–111.

Ho, J.-Y., & O'Sullivan, E. (2017). Strategic standardisation of smart systems: A roadmapping process in support of innovation. *Technological Forecasting and Social Change*, *115*, 301–312.

Ho, J.-Y., & O'Sullivan, E. (2018). Standardisation framework to enable complex technological innovations: The case of photovoltaic technology. *Journal of Engineering and Technology Management*, *50*, 2–23.

Hobday, A. J., Boschetti, F., Moeseneder, C., Stephenson, R. L., Bessey, C., Bulman, C. M., ... others (2020). Quantitative foresighting as a means of improving anticipatory scientific capacity and strategic planning. *One Earth*, *3*(5), 631–644.

Højland, J., & Rohrbeck, R. (2018). The role of corporate foresight in exploring new markets–evidence from 3 case studies in the bop markets. *Technology Analysis & Strategic Management*, *30*(6), 734–746.

Iden, J., Methlie, L. B., & Christensen, G. E. (2017). The nature of strategic foresight research: A systematic literature review. *Technological Forecasting and Social Change*, *116*, 87–97.

Idoko, O., & MacKay, R. B. (2021). The performativity of strategic foresight tools: Horizon scanning as an activation device in strategy formation within a uk financial institution. *Technological Forecasting and Social Change*, *162*, 120389.

Ilmola, L., & Kuusi, O. (2006). Filters of weak signals hinder foresight: Monitoring weak signals efficiently in corporate decision-making. *Futures*, *38*(8), 908–924.

Iris.ai. (2022). *Literature review tools.* (Retrieved from: https://iris.ai/features/ (06.01.2022))

Klos, C., & Spieth, P. (2021). Ready, steady, digital?! how foresight activities do (not) affect individual technological frames for managerial sensemaking. *Technological Forecasting and Social Change*, *163*, 120428.

Lehr, T., Lorenz, U., Willert, M., & Rohrbeck, R. (2017). Scenario-based strategizing: Advancing the applicability in strategists' teams. *Technological Forecasting and Social Change*, *124*, 214–224.

Li, A., & Sullivan, B. N. (2022). Blind to the future: Exploring the contingent effect of managerial hubris on strategic foresight. *Strategic Organization*, *20*(3), 565–599.

Luzinski, C. (2014). Identifying leadership competencies of the future: Introducing the use of strategic foresight. *Nurse Leader*, *12*(4), 37–47.

Maertins, A. (2016). From the perspective of capability: Identifying six roles for a successful strategic foresight process. *Strategic Change*, *25*(3), 223–237.

Major, E., & Cordey-Hayes, M. (2000). Engaging the business support network to give smes the benefit of foresight. *Technovation*, *20*(11), 589–602.

Marinković, M., Al-Tabbaa, O., Khan, Z., & Wu, J. (2022). Corporate foresight: A systematic literature review and future research trajectories. *Journal of Business Research*, *144*, 289–311.

Martin, B. R. (2010). The origins of the concept of 'foresight'in science and technology: An insider's perspective. *Technological Forecasting and Social Change*, *77*(9), 1438–1447.

Mastio, E., & Dovey, K. (2021). Contextual insight as an antecedent to strategic foresight. *Futures*, *128*, 102715.

Miller, R. (2018). Sensing and making-sense of futures literacy: towards a futures literacy framework (flf). In *Transforming the future* (pp. 15–50). Routledge.

Miller, R., & Sandford, R. (2019). Futures literacy: The capacity to diversify conscious human anticipation. *Handbook of anticipation*, 73–91.

Milshina, Y., & Vishnevskiy, K. (2018). Potentials of collaborative foresight for SMEs. *Technology Analysis & Strategic Management*, *30*(6), 701–717.

Moore, J. (2018). Hands-on or heads-up? strategic foresight as the heart of the board's work. *Board Leadership*, *2018*(159), 7–8.

Mühlroth, C., & Grottke, M. (2018). A systematic literature review of mining weak signals and trends for corporate foresight. *Journal of Business Economics*, *88*(5), 643–687.

Nkuda, M. O. (2017). Strategic agility and competitive advantage: Exploration of the ontological, epistemological and theoretical underpinnings. *British Journal of Economics, Management & Trade*, *16*(1), 1–13.

Nugroho, Y., & Saritas, O. (2009). Incorporating network perspectives in foresight: A methodological proposal. *Foresight*.

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying information technology in organizations: Research approaches and assumptions. *Information systems research*, *2*(1), 1–28.

Paliokaitė, A., & Pačėsa, N. (2015). The relationship between organisational foresight and organisational ambidexterity. *Technological Forecasting and Social Change*, *101*, 165–181.

Paliokaitė, A., Pačėsa, N., & Sarpong, D. (2014). Conceptualizing strategic foresight: An integrated framework. *Strategic change*, *23*(3-4), 161–169.

Peter, M. K., & Jarratt, D. G. (2015). The practice of foresight in long-term planning. *Technological Forecasting and Social Change*, *101*, 49–61.

Peterson, A., & Wu, A. (2021). Entrepreneurial learning and strategic foresight. *Strategic Management Journal*, *42*(13), 2357–2388.

Piirainen, K. A., & Gonzalez, R. A. (2015). Theory of and within foresight—"what does a theory of foresight even mean?". *Technological Forecasting and Social Change*, *96*, 191–201.

Pulsiri, N., & Vatananan-Thesenvitz, R. (2021). Triangle relationship: a review of dynamic capabilities, strategic foresight and organizational learning. *International Journal of Business Environment*, *5*(3), 1–18.

Ramírez, R., Österman, R., & Grönquist, D. (2013). Scenarios and early warnings as dynamic capabilities to frame managerial attention. *Technological Forecasting and Social Change*, *80*(4), 825–838.

Rasmussen, B., Andersen, P. D., & Borch, K. (2010). Managing transdisciplinarity in strategic foresight. *Creativity and Innovation Management*, *19*(1), 37–46.

Reid, D. M., & Zyglidopoulos, S. C. (2004). Causes and consequences of the lack of strategic foresight in the decisions of multinational enterprises to enter china. *Futures*, *36*(2), 237–252.

Reimers-Hild, C. (2018). Strategic foresight, leadership, and the future of rural healthcare staffing in the united states. *Jaapa*, *31*(5), 44.

Rhisiart, M., Miller, R., & Brooks, S. (2015). Learning to use the future: developing foresight capabilities through scenario processes. *Technological Forecasting and Social Change*, *101*, 124–133.

Rincón, G. B., & Díaz-Domínguez, A. (2022). Assessing futures literacy as an academic competence for the deployment of foresight competencies. *Futures*, *135*, 102872.

Ringland, G. (2010). The role of scenarios in strategic foresight. *Technological Forecasting and Social Change*, *77*(9), 1493–1498.

Rohrbeck, R. (2010). *Corporate foresight: towards a maturity model for the future orientation of a firm*. Springer Science & Business Media.

Rohrbeck, R. (2012). Exploring value creation from corporate-foresight activities. *Futures*, *44*(5), 440–452.

Rohrbeck, R., Battistella, C., & Huizingh, E. (2015). Corporate foresight: An emerging field with a rich tradition. *Technological Forecasting and Social Change*, *101*, 1–9.

Rohrbeck, R., & Gemünden, H. G. (2011). Corporate foresight: Its three roles in enhancing the innovation capacity of a firm. *Technological forecasting and social change*, *78*(2), 231–243.

Rohrbeck, R., & Kum, M. E. (2018). Corporate foresight and its impact on firm performance: A longitudinal analysis. *Technological Forecasting and Social Change*, *129*, 105–116.

Rohrbeck, R., & Schwarz, J. O. (2013). The value contribution of strategic foresight: Insights from an empirical study of large european companies. *Technological Forecasting and Social Change*, *80*(8), 1593–1606.

Rohrbeck, R., Thom, N., & Arnold, H. (2015). It tools for foresight: The integrated insight and response system of deutsche telekom innovation laboratories. *Technological Forecasting and Social Change*, *97*, 115–126.

Ruff, F. (2006). Corporate foresight: integrating the future business environment into innovation and strategy. *International Journal of Technology Management*, *34*(3-4), 278–295.

Ruff, F. (2015). The advanced role of corporate foresight in innovation and strategic management—reflections on practical experiences from the automotive industry. *Technological Forecasting and Social Change*, *101*, 37–48.

Sarpong, D., & Hartman, D. (2018). Fading memories of the future: the dissipation of strategic foresight among middle managers. *Technology Analysis & Strategic Management*, *30*(6), 672–683.

Sarpong, D., & Maclean, M. (2016). Cultivating strategic foresight in practise: A relational perspective. *Journal of Business Research*, *69*(8), 2812–2820.

Sarpong, D., Maclean, M., & Alexander, E. (2013). Organizing strategic foresight: A contextual practice of 'way finding'. *Futures*, *53*, 33–41.

Sarpong, D., Maclean, M., & Davies, C. (2013). A matter of foresight: How practices enable (or impede) organizational foresightfulness. *European Management Journal*, *31*(6), 613–625.

Savioz, P., & Blum, M. (2002). Strategic forecast tool for smes: how the opportunity landscape interacts with business strategy to anticipate technological trends. *Technovation*, *22*(2), 91–100.

Scheiner, C. W., Baccarella, C. V., Bessant, J., & Voigt, K.-I. (2015). Thinking patterns and gut feeling in technology identification and evaluation. *Technological Forecasting and Social Change*, *101*, 112–123.

Schoemaker, P. J., & Tetlock, P. E. (2017). Building a more intelligent enterprise. *MIT Sloan Management Review*.

Schwarz, J. O., Ram, C., & Rohrbeck, R. (2019). Combining scenario planning and business wargaming to better anticipate future competitive dynamics. *Futures*, *105*, 133–142.

Schweitzer, N., Hofmann, R., & Meinheit, A. (2019). Strategic customer foresight: From research to strategic decision-making using the example of highly automated vehicles. *Technological Forecasting and Social Change*, *144*, 49–65.

Singh, S., Dhir, S., Das, V. M., & Sharma, A. (2020). Bibliometric overview of the technological forecasting and social change journal: Analysis from 1970 to 2018. *Technological Forecasting and Social Change*, *154*, 119963.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of business research*, *104*, 333–339.

Tapinos, E. (2013). Scenario planning at business unit level. *Futures*, *47*, 17–27.

Tapinos, E., & Pyper, N. (2018). Forward looking analysis: Investigating how individuals 'do'foresight and make sense of the future. *Technological Forecasting and Social Change*, *126*, 292–302.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic management journal*, *18*(7), 509–533.

Un, S., & Price, N. (2007). Bridging the gap between technological possibilities and people:: Involving people in the early phases of technology development. *Technological Forecasting and Social Change*, *74*(9), 1758–1772.

Van der Duin, P., Heger, T., & Schlesinger, M. D. (2014). Toward networked foresight? exploring the use of futures research in innovation networks. *Futures*, *59*, 62–78.

van der Laan, L. (2021). Disentangling strategic foresight? a critical analysis of the term building on the pioneering work of richard slaughter. *Futures*, *132*, 102782.

van der Laan, L., & Erwee, R. (2012). Foresight styles assessment: a valid and reliable measure of dimensions of foresight competence? *Foresight*.

Vecchiato, R. (2012). Strategic foresight: matching environmental uncertainty. *Technology Analysis & Strategic Management*, *24*(8), 783–796.

Vecchiato, R. (2015). Creating value through foresight: First mover advantages and strategic agility. *Technological Forecasting and Social Change*, *101*, 25–36.

Vecchiato, R. (2020). Analogical reasoning, cognition, and the response to technological change: lessons from mobile communication. *Research Policy*, *49*(5), 103958.

Vecchiato, R., & Roveda, C. (2010a). Foresight in corporate organisations. *Technology Analysis & Strategic Management*, *22*(1), 99–112.

Vecchiato, R., & Roveda, C. (2010b). Strategic foresight in corporate organizations: Handling the effect and response uncertainty of technology and social drivers of change. *Technological Forecasting and Social Change*, *77*(9), 1527–1539.

Von der Gracht, H. A., Bañuls, V. A., Turoff, M., Skulimowski, A. M., & Gordon, T. J. (2015). *Foresight support systems: The future role of ict for foresight* (Vol. 97). Elsevier.

Von der Gracht, H. A., & Stillings, C. (2013). An innovation-focused scenario process—a case from the materials producing industry. *Technological Forecasting and Social Change*, *80*(4), 599–610.

Von Der Gracht, H. A., Vennemann, C. R., & Darkow, I.-L. (2010). Corporate foresight and innovation management: A portfolio-approach in evaluating organizational development. *Futures*, *42*(4), 380–393.

Weber, C., Sailer, K., & Katzy, B. (2015). Real-time foresight—preparedness for dynamic networks. *Technological Forecasting and Social Change*, *101*, 299–313.

Weigand, K., Flanagan, T., Dye, K., & Jones, P. (2014). Collaborative foresight: Complementing long-horizon strategic planning. *Technological Forecasting and Social Change*, *85*, 134–152.

Wiener, M., Gattringer, R., & Strehl, F. (2018). Participation in interorganisational collaborative open foresight a matter of culture. *Technology Analysis & Strategic Management*, *30*(6), 684–700.

Wiener, M., Gattringer, R., & Strehl, F. (2020). Collaborative open foresight-a new approach for inspiring discontinuous and sustainability-oriented innovations. *Technological Forecasting and Social Change*, *155*, 119370.

Wright, G., O'Brien, F., Meadows, M., Tapinos, E., & Pyper, N. (2020). *Scenario planning and foresight: Advancing theory and improving practice* (Vol. 159). Elsevier.

Wright, G., Van der Heijden, K., Burt, G., Bradfield, R., & Cairns, G. (2008). Scenario planning interventions in organizations: An analysis of the causes of success and failure. *Futures*, *40*(3), 218–236.

Yoon, J., Kim, Y., Vonortas, N. S., & Han, S. W. (2018). Corporate foresight and innovation: the effects of integrative capabilities and organisational learning. *Technology Analysis & Strategic Management*, *30*(6), 633–645.

Yoon, J., Kim, Y. J., Vonortas, N. S., & Han, S. W. (2019). A moderated mediation model of technology roadmapping and innovation: The roles of corporate foresight and organizational support. *Journal of Engineering and Technology Management*, *52*, 61–73.

Junior Management Science 8(3) (2023) 682-716

Junior Management Science

journal homepage: www.jums.academy

Turning German Steel Production Green: Quantifying Diffusion Scenarios for Hydrogen-Based Steelmaking and Policy Implications

Philipp Preis

Technische Universität München

Abstract

The German steel industry is in jeopardy. Current steel production must be comprehensively transformed to achieve the emission targets imposed by the Federal Climate Change Act. A promising alternative that has increasingly gained momentum in recent years is hydrogen-based steel production. This thesis analyzes the potential of this method to transform the German steel industry. First, drivers that will decisively influence the future role of hydrogen-based steelmaking are identified. Subsequently, these drivers are linked in a quantitative model to develop explorative diffusion scenarios and to draw conclusions for policymaking. Four representative scenarios are extracted and analyzed. Large differences between the scenario outputs illustrate that the diffusion of hydrogen-based steelmaking is subject to significant uncertainties. It becomes clear that the most effective lever for promoting the attractiveness of hydrogen-based steelmaking is increasing the cost of conventional production by exposing it to CO_2 prices. However, such exposure simultaneously suggests disadvantages towards producers that are not subject to this regulation. To mitigate the emerging risk of carbon leakage effects, suitable policy measures are required.

Keywords: Green steel; Green hydrogen; Energy transition; Energy policy.

1. Introduction

"It is unequivocal that human influence has warmed the atmosphere, ocean and land. Widespread and rapid changes in the atmosphere, ocean, cryosphere and biosphere have occurred."[1] With these words, the Intergovernmental Panel on Climate Change introduces the first part of its Sixth Assessment Report, highlighting the impact of anthropogenic greenhouse gas emissions on the global climate to date. Accordingly, any further rise in global temperature is expected to increase the likelihood of drastic consequences such as heat waves, droughts, floods, or extreme weather events to occur.[2]

To mitigate such consequences, the Paris Agreement was adopted at the United Nations Climate Change Conference in 2015. In this agreement, the attending parties committed to limiting global warming to below 2°C compared to pre-industrial levels.[3] However, deep reductions in global greenhouse gas emissions are required to achieve this target.[4]

In response, German policymakers have updated the national climate targets of Germany and set the goal of becoming carbon-neutral by 2045.[5] These targets pose major challenges for many industries as their processes must be adapted to the new objectives. One industry that is significantly affected is the German steel production. In 2019, this industry was responsible for almost 7% of Germany's total emissions, generating nearly 25 times the emissions of national aviation.[6] Due to this relevance for the emissions balance of the whole economy, steel producers are increasingly under pressure. Current steelmaking must be thoroughly decarbonized to align it with the climate targets and ensure the long-term preservation of the industry.[7]

[1] IPCC, 2021, p. SPM5.

[2] Cf. IPCC, 2021, pp. SPM21-SPM25.

[3] Cf. UN, 2015, Article 2.

[4] Cf. IPCC, 2021, p. SPM17.

[5] Cf. Federal Climate Change Act, Section 3.

[6] Including emissions from energy installations, cf. DEHst (2014-2021), pp. 26-27; UBA, 2022.

[7] Cf. Fischedick, Marzinkowski, Winzer, & Weigel, 2014, p. 574; Kushnir, Hansen, Vogl, & Åhman, 2020, p. 12; Müller, Herz, Reichelt, Jahn, & Michaelis, 2021, p. 2; Vogl, Åhman, & Nilsson, 2021, p. 79.

DOI: https://doi.org/10.5282/jums/v8i3pp682-716

In light of these challenges, hydrogen-based steel production has gained momentum in recent years. The basic idea behind this method is to achieve enormous emission savings by utilizing hydrogen produced through renewable energies as reduction agent.[8] Although hydrogen-based steelmaking offers great potential, its introduction is still subject to significant uncertainties. For instance, hydrogen-based steelmaking is currently limited to research and development projects beyond the stage of large-scale market readiness.[9] Furthermore, higher production costs compared to conventional production are expected.[10] Problems like these indicate that extensive political support is required to achieve a sustainable and timely transformation of the German steel industry.[11]

The motivation underlying this thesis is derived from the described problem set. The first objective is to assess the current framework for introducing hydrogen-based steel production and identify drivers that will most significantly influence its diffusion. The identified drivers will then be linked in a quantitative model to develop explorative scenarios for the diffusion of hydrogen-based steel production in Germany. Based on the developed scenarios, foreseeable developments and key interrelationships in the defined environment will subsequently be derived. From these findings, the aim is to draw conclusions for policymakers regarding targeted support of German steel production. The following research questions reflect the key objectives of this thesis:

1. Which drivers will decisively shape the diffusion of hydrogen-based steel production in Germany?
2. Based on these drivers, which key scenarios can be derived for the diffusion of hydrogen-based steel production in Germany?
3. How should policymakers act to foster the diffusion of hydrogen-based steel production in Germany sustainably?

First, a brief overview of previous modeling approaches is provided in Chapter 2 to derive the additional informative value of this thesis. Chapter 3 then analyzes the environment of the German steel industry from various perspectives to gain initial insights into the first research question and to create a qualitative foundation for the subsequent model development. The model development then takes place in Chapter 4: After describing the underlying methodology, suitable scenarios are extracted, evaluated, and discussed in the context of the research questions. The conclusion, as well as an outlook on future research potential, are provided in Chapter 5.

2. Literature Research

In this chapter, excerpts of previous research are presented in order to derive the scope of this thesis. First, studies that investigated the overall potential of hydrogen-based steel production on the technological level are considered. Secondly, approaches that provide a perspective on the potential of hydrogen-based steelmaking in a systematic application are examined. Lastly, a review of previous work regarding policy implications is provided.

At the technological level, the assessment of hydrogen-based steel production has already been part of several studies. Fischedick et al. (2014) compared alternative technologies within a techno-economic model. They concluded that hydrogen-based steelmaking, also known as hydrogen direct reduction (H-DR), will only show sufficient profitability for actual introduction between 2030 and 2040.[12] Vogl, Åhman, and Nilsson (2018) also investigated the H-DR method in terms of its competitiveness against conventional steel production and deduced that it is fundamentally associated with higher costs, which are highly dependent on specific factors. Furthermore, the authors consider the H-DR method as an option to achieve the long-term emission targets of the European Union.[13] Jacobasch et al. (2021) predicted that H-DR production will have lower production costs as well as ecological advantages over conventional production by 2050.[14]

In a systematic context, an early approach is provided by Woertler et al. (2013). The authors considered various production methods in the framework of the entire European steel industry and with respect to the European climate targets. They concluded, among others, that saving about 10% of the steel industry's 1990 emissions is the most likely scenario and will be achieved mainly by improving current processes.[15] Kushnir et al. (2020) analyzed systematic conditions within Sweden to assess the potential for a switch to hydrogen-based steelmaking. The authors characterized H-DR production as the best available option to meet Swedish climate targets but derived major barriers and the need for strong policy support.[16] Similarly, in the context of the Swedish steel industry, Toktarova et al. (2020) developed a model to analyze specific pathways to achieve deep emission reductions. These pathways differ, for example, in applied production methods and different assumptions of steel output development. The authors concluded that establishing H-DR production offers significant abatement potential but is associated with challenges due to its high electricity consumption.[17] A similar approach is provided by Arens, Worrell, Eichhammer, Hasanbeigi, and Zhang (2017) for the German steel industry. They, too, defined individual pathways to analyze the emission reduction potential until 2035. Their approach focused on the emission abatement potential of the pathways and did not consider the production costs of the individual methods. In this analysis, the authors found that the European emission targets for 2030 can only be achieved through substantial reductions in production vol-

[8] Cf. Otto et al., 2017, p. 10.
[9] Cf. Kushnir et al., 2020, p. 2.
[10] Cf. BMWI, 2020b, p. 15.
[11] Cf. Kushnir et al., 2020, p. 12; Vogl et al., 2021, p. 79.
[12] Cf. Fischedick et al., 2014, p. 563.
[13] Cf. Vogl et al., 2018, p. 744.
[14] Cf. Jacobasch et al., 2021, p. 18.
[15] Cf. Woertler et al., 2013, p. 5.
[16] Cf. Kushnir et al., 2020, p. 12.
[17] Cf. Toktarova et al., 2020, pp. 14-15.

ume.[18]

Policy-based analyses and detailed recommendations for promoting sustainable production methods currently exist mainly at superordinate levels, such as basic materials or energy-intensive industries.[19] In the context of steel production, several studies identified the need for policy support for transforming the industry but do not offer specific approaches or recommendations.[20] More detailed results are provided by Vogl et al. (2021). The researchers analyzed different policy approaches for promoting the early market introduction phase of sustainable steel and derived potential especially in direct subsidies for steel production.[21] Furthermore, Muslemani, Liang, Kaesehage, Ascui, and Wilson (2021) investigated the potential to promote green steel and products thereof by creating separate markets. The authors concluded that policy approaches would be particularly promising if these included measures that consider potentially emerging distortions of competition across countries and sectors.[22]

Relating these results to the future of H-DR production in Germany, many ambiguities arise, which previous research has not answered. For instance, uncertainties exist about how exactly the development of production costs or other essential factors might affect the diffusion of H-DR production. Furthermore, the current policy regulations and targets suggest playing a critical role in the steel industry's future. However, its systematic implications on the German steel industry were rarely analyzed in detail. This raises questions such as to what extent the current emission targets are compatible with the steel industry in the short and long term and which specific levers could be used by policymakers to exert influence effectively. This thesis contributes to the clarification of these and other questions.

3. Analysis of the Initial Situation

This section forms the qualitative foundation for the subsequent quantitative scenario development regarding the diffusion of hydrogen-based steel production in Germany. Therefore, this chapter aims to define the initial situation and to identify the major challenges as well as the most significant drivers influencing this diffusion. These findings will then be utilized to draw a plausible picture of the current and foreseeable framework conditions as inputs for the scenario development. For this purpose, the German steel industry will first be characterized with a focus on its economic setting, followed by an analysis of the prevailing environment from various perspectives. These are divided into technological, industry-specific, and political aspects, with a particular emphasis on factors most likely to influence the adoption of more sustainable methods and especially H-DR steelmaking.

3.1. Profile of the German Steel Industry and Its Economic Environment

Generally, iron and steel production, like many other industrial sectors, is characterized by increased difficulty of decarbonization. The reasons for this can be found in the long lifetimes of production plants and corresponding infrastructure, as well as the lack of less emission-intensive alternative technologies.[23] Furthermore, the heterogeneity of industrial plants and the frequent utilization of fossil fuels as input material increasingly complicate decarbonization.[24] Additionally, the steel industry is attributed to the hard-to-abate sectors, typified by a non-electric supply of their energy requirements and difficult or even impossible electrification due to reasons like high costs or technical barriers.[25]

These complications also become evident when considering the German steel industry. In terms of emission intensity, no discernible progress has been observed in recent years, as the overall emission intensity remained relatively constant. In 2013, an average of 1.34 metric tons of carbon dioxide equivalents (tCO_2-eq) per metric ton of steel (tSteel) were generated, while in 2020, this figure had slightly risen to 1.35 tCO_2-eq/tSteel.[26] This immense intensity led to emissions of about 48.2 million metric tons (MMT) of CO_2-eq in 2020, corresponding to 6.6% of total German emissions.[27]

In Germany, steel production is a core industry with a long history. It has about 83,000 direct employees and is closely linked to other major sectors such as automotive, mechanical engineering, and construction.[28] In 2020, 35.7 MMT of steel were produced, representing an exceptional drop compared to the 39.6 MMT produced in 2019, mainly attributable to impacts caused by the Covid-19 pandemic. This decline became even more evident in the generated revenue, which dropped by over 19% between 2019 and 2020, from € 39.8 billion to € 32.1 billion.[29] However, a quick recovery from the crisis can already be observed. In 2021, the total steel production increased to about 40.1 MMT of steel, exceeding the production level of 2019.[30] In an international context, this production volume makes the German steel industry the largest in the European Union, representing a world market share of 2.1% in terms of crude steel produced in 2021.[31] A large part of this total production is accounted for by single players dominating the market. The three biggest market

[18] Cf. Arens et al., 2017, p. 89.
[19] Cf. Nilsson et al., 2021; Sartor & Bataille, 2019; Wyns, Khandekar, Axelson, Sartor, & Neuhoff, 2019.
[20] Cf. Fan & Friedmann, 2021, p. 856; Holappa, 2020, p. 15; Weigel, Fischedick, Marzinkowski, & Winzer, 2016, p. 1074.
[21] Cf. Vogl et al., 2021, p. 78.
[22] Cf. Muslemani et al., 2021, pp. 10-11.
[23] Cf. IEA, 2021, p. 135.
[24] Cf. Bhaskar, Assadi, & Nikpey Somehsaraei, 2020, p. 1.
[25] Cf. IEA, 2019, p. 23.
[26] Including emissions from energy installations, cf. DEHst, 2014-2021; Worldsteel, 2009-2022, p. 1.
[27] Emissions of the German steel industry consisted of 16.8 MMT CO_2-eq from own energy installations and 31.4 MMT CO_2-eq from process emissions, cf. DEHst, 2014-2021, p. 28. Total German emissions in 2020 amounted to 729 MMT CO_2-eq, cf. UBA, 2022.
[28] Cf. WV Stahl, 2021b, pp. 11-12.
[29] Cf. WV Stahl, 2021b, pp. 7, 13.
[30] Cf. WV Stahl, 2022, p. 1.
[31] Cf. Worldsteel, 2009-2022.

players are thyssenkrupp, ArcelorMittal and Salzgitter AG. In 2020, these were responsible for the production of around 11.0, 6.5, and 6.0 MMT of crude steel, respectively, and thus accounted for roughly 66% of total steel production in Germany.[32]

A distinction is drawn between two types of steel: primary and secondary steel. Primary steel is produced from virgin iron ore and is usually of high quality. For this reason, it is mainly used for the production of flat steel products for application in industries such as automotive or machine building. Secondary steel is produced by recycling steel scrap, which results in inferior quality. Therefore, it is mainly employed to create long steel products for applications in construction.[33] German steel producers focus on producing high-quality primary steel, which is reflected, for example, in a consistent export surplus of steel scrap.[34] As shown in Figure 1, primary steel production thus accounts for around 70% of the total production volume, while secondary steel production accordingly accounts for around 30%.

From an economic perspective, the German steel industry is facing increasing difficulties. The consideration of key indicators regarding its economic situation highlights that these have often been mediocre or even declining in recent years. The volume of crude steel produced, and the sales revenues generated show a negative growth path between 2010 and 2019, with compound annual growth rates of -1.1% and -0.3%.[36] During the financial crisis in 2009 and the Covid-19 pandemic in 2020, the German steel industry reacted sensitively. During both, German steel production slumped sharply compared to global levels.[37]

Initial explanatory approaches for this development can be found in fundamental characteristics of the steel industry, which complicate operations. As such, the production of steel is facing high entry barriers. Furthermore, characteristics of energy-intensive industries like long investment cycles and high capital intensity discourage investors, reducing the economic attractiveness of the industry.[38] However, the most striking reason for this development is provided in the challenging market environment of the German steel industry. In many countries, enormous overcapacities exist, which seriously impair the functioning of global steel markets. In addition, protectionist measures by trading partners are weakening exports, and competition from subsidized manufacturers offering steel at significantly lower prices is increasingly distorting competition.[39] In particular, the Chinese steel industry strongly influenced global steel production in recent years and increased its world market share from 15% in 2000 to over 53% in 2019.[40]

Additionally, steel production in Europe and Germany is characterized as cost-intensive compared to other countries, which further impedes the globally competitive pricing of German steel. These pricing constraints are reflected in an analysis of global steel production costs by the Joint Research Centre of the European Commission. The authors concluded that European and thus German manufacturers are among the most expensive steel producers globally, especially in terms of raw material and labor costs.[41] As shown in Figure 1, in their entirety, these factors have caused the global market share of German steel production to drop by 1.3 percentage points since 2008, despite an increase in the global market of around 40% during the same period in terms of production volume.[42]

Thus, it can be concluded that the German steel industry finds itself in a difficult economic situation to implement and finance a large-scale transformation of current production capacities. In the context of the transformation towards hydrogen-based production, the latter in particular appears to be a major challenge: The German Steel Federation, representing the political interests of German steel producers, estimates that the transition of German primary steel production would require investments of around € 30 billion, almost as high as total industry sales in 2020, and hence derives significant burdens for steel producers.[43]

3.2. Technological Environment

In order to gain a more precise understanding of the initial situation, it is essential to consider current as well as foreseeable technological circumstances of steel production. For this purpose, an analysis of these processes is conducted, followed by the identification of promising alternatives from literature and a characterization of hydrogen-based steel production.

3.2.1. Currently Applied Production Technologies

At present, primary and secondary steel production each takes place within the framework of one dominant production method. These methods will be explained in the following.

In primary steel production, the raw materials in the form of coal and iron ore must initially be processed separately in a sintering or coking plant to obtain the intermediate products sinter and coke. Sinter consists of small lumps produced by melting the iron ore (Fe2O3). Coke (C) serves as energy source and is produced by heating coal to remove volatile fractions. These are then added to a blast furnace (BF) along with lime fluxes, which are used to control the impurity level

[32] Cf. ArcelorMittal, 2021, p. 74; Salzgitter AG, 2021, p. 2; Thyssenkrupp, 2021, p. 68.

[33] Cf. Arens, Åhman, & Vogl, 2021, p. 4; Woertler et al., 2013, pp. 6-8.

[34] Cf. Arens et al., 2017, p. 86; WV Stahl, 2021a, p. 3.

[35] Cf. Worldsteel, 2009-2022; WV Stahl, 2022. Due to lack of data, 2019 primary/ secondary split adopted for 2020.

[36] Cf. WV Stahl, 2021b, pp. 7, 13.

[37] 2009 global/ German growth: -8%/ -29%; 2020: 0%/ -10%. Cf. Worldsteel, 2009-2022.

[38] Cf. Karakaya, Nuur, & Assbring, 2018, p. 651; Wesseling et al., 2017, p. 1311.

[39] Cf. EC, 2021b, pp. 5, 23.

[40] Cf. Worldsteel, 2009-2022.

[41] Cf. Medarac, Moya, & Somers, 2020, p. 15.

[42] Global steel production increased from 1,343 MMT in 2008 to 1,875 MMT in 2019, cf. Worldsteel, 2009-2022.

[43] Cf.WV Stahl, 2021c, p. 4.

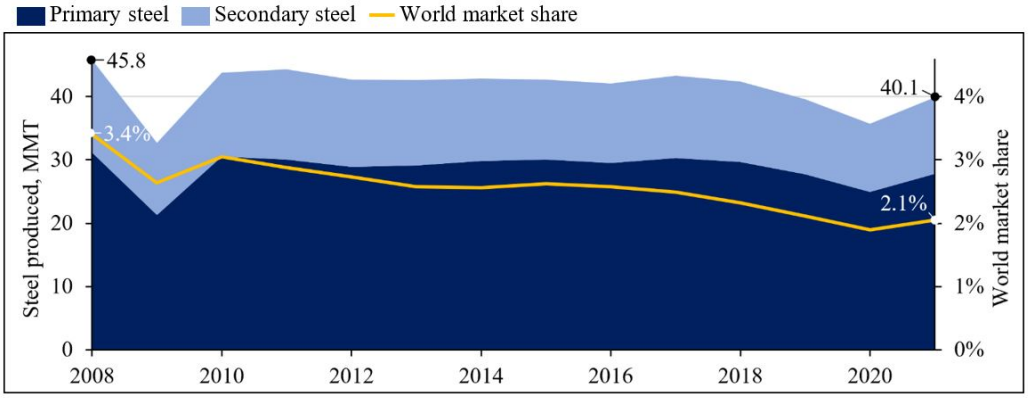


Figure 1: German steel production volume (2008-2021).[35]

and temperature. This mixture is called the burden.[44] Iron production then takes place in the BF by passing a stream of hot air, pulverized coal, and oxygen through it. Iron (Fe) is produced through the reduction of the iron ore by the coke or carbon monoxide (CO) with the simultaneous formation of carbon dioxide (CO_2), as illustrated in the following equations:[45]

$$Fe_2O_3 + \frac{3}{2}C \rightarrow 2Fe + \frac{3}{2}CO_2 \quad (1)$$

$$Fe_2O_3 + 3CO \rightarrow 2Fe + 3CO_2 \quad (2)$$

The product of this step is pig iron, which contains too much carbon resulting in increased brittleness of the material. To reduce the carbon content from around 4% to 0.25%, the liquid iron must be heated again in a basic oxygen furnace (BOF) while adding steel scrap and oxygen, a process referred to as steelmaking. The steel can then be further processed through additional steps such as casting or rolling to produce the required final product.[46] Due to the combination of a BF with a BOF present, this production route is also referred to as BF-BOF. The production steps and major material flows are summarized in Figure 2.

Secondary steel production is less complex. This process is based on recycling steel scrap, which is melted in just one step by heating it with electrical energy in an electric arc furnace (EAF). Additives such as coal and natural gas as complementary energy sources as well as lime fluxes are required. Furthermore, electrodes are consumed.[48] This production route is summarized in Figure 3.

Although the secondary production route only provides steel of limited quality, it incorporates a significant advantage in its comparatively low emission intensity. While emissions from primary production are usually reported at 1.7-1.9 tCO_2-eq/tSteel,[50] emissions from the secondary route are much lower at around 0.3-0.5 tCO_2-eq/tSteel.[51] Since only about 0.1 tCO_2-eq/tSteel of these consist of direct emissions,[52] the majority result from indirect emissions caused by the emission intensity of the respective power grid. These could thus be eliminated by decarbonizing electricity generation. The prospects for the integrated route are much worse: The theoretical minimum, determined by chemical limitations, is 1.37 tCO_2-eq/tSteel, a multiple of the already realized emissions within the recycled route.[53]

This analysis shows that primary steel production in particular is responsible for the largest part of emissions. Not only does it cause up to six times more emissions per metric ton of steel than the secondary route, but it is also utilized to produce around 70% of the total steel volume, thus causing over 90% of all emissions of German steel production.[54] As the prospects for potential emission savings are also limited, this displays a high degree of incompatibility with decarbonization efforts. For this reason, decarbonization of the steel industry by substituting BF-BOF production has been a focal topic of discussion in literature for years and will be dealt with in greater depth in the next section.

3.2.2. Alternatives for Currently Applied Production Technologies

Consensus exists that large-scale decarbonization of steel production can only be realized by comprehensively transforming current steel production.[55] However, other ap-

[44] Cf. Bailera, Lisbona, Peña, & Romeo, 2021, p. 3; IEA, 2020, pp. 27-29.
[45] Cf. Bailera et al., 2021, pp. 3-4; Otto et al., 2017, pp. 5-6.
[46] Cf. Birat, 2020, p. 6; IEA, 2020, pp. 19, 29; Otto et al., 2017, p. 6.
[47] Own illustration based on process description above.
[48] Cf. Demus, Reichel, Schulten, Echterhof, & Pfeifer, 2016, p. 565; Otto et al., 2017, p. 7.
[49] Own illustration based on process description above.
[50] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 164; Bhaskar et al., 2020, p. 2; Germeshuizen & Blom, 2013, p. 10673; Vogl et al., 2018, p. 740; Weigel et al., 2016, p. 568.
[51] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 51; Kirschen, Badr, & Pfeifer, 2011, p. 6148; Morfeldt, Nijs, & Silveira, 2015, p. 2.
[52] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 52; Demus et al., 2016, p. 565.
[53] Cf. Schoemaker, 1995 (as qtd. in Kirschen et al., 2011, p. 6148).
[54] When considering 1.8 tCO_2-eq/tSteel for BF-BOF and 0.4 tCO_2-eq/tSteel for EAF.
[55] Cf. Fischedick et al., 2014, p. 574; Kushnir et al., 2020, p. 12; Müller et al., 2021, p. 2; Vogl et al., 2021, p. 79.

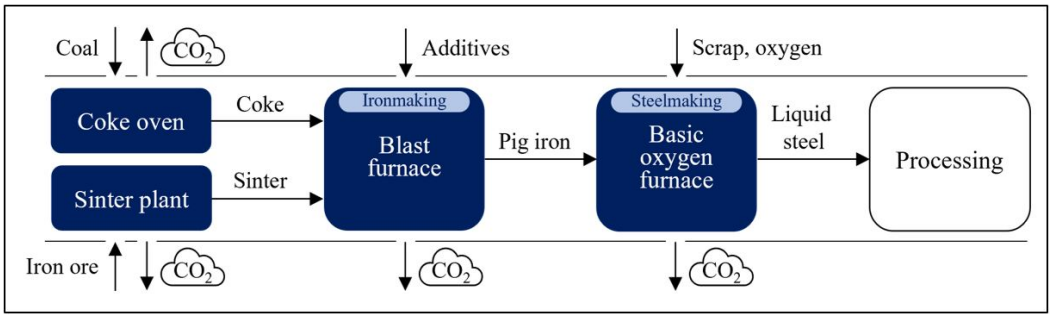


Figure 2: Primary steel production within the BF-BOF production route.[47]

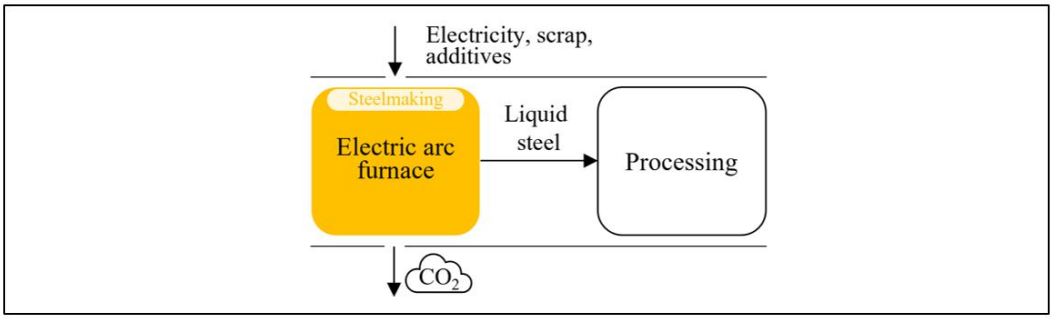


Figure 3: Secondary steel production within the EAF production route.[49]

proaches apart from hydrogen-based technologies exist, of which an overview is provided below.

A variety of studies have already investigated energy and material efficiency strategies for reducing emissions and have derived considerable potential from material efficiency strategies in particular.[56] Another area of research is the development of secondary steel production in the EAF. Pauliuk, Milford, Müller, and Allwood (2013) concluded that secondary steel production will double globally by 2050, replacing primary production as the dominant route between 2050 and 2060.[57] Xylia, Silveira, Duerinck, and Meinke-Hubeny (2018) presented similar results, predicting that the share of secondary production will be around 50% in 2050 and will become the globally dominant route by 2060.[58] Further improvements to this production route are also in prospect: Reducing direct emissions from the EAF could be realized by substituting the applied fossil fuels with biological alternatives such as biochar, which would enable an entirely carbon-neutral secondary steel production.[59]

Nevertheless, these results suggest that primary steel production will still be required in the coming decades, as Vogl et al. (2021) even conclude for the "(...) most ambitious circular economy scenarios (...)"[60] in Europe. Various studies have investigated alternatives to realize emission savings through incremental or radical technology shifts. For example, the operation of conventional BF-BOF production with additional recycling of the furnace gas or the application of carbon capture and storage (CCS) technologies were considered.[61] Similarly, the use of bioenergy as energy source could reduce total emissions by up to 20%, as concluded by Mandova et al. (2019).[62] More radical solutions are found in novel methods such as the electrolysis of iron oxide or the electrification of production within the range of various power-to-X processes.[63] As Weigel et al. (2016) determined in the course of a multi-criteria analysis and Jacobasch et al. (2021) via an economic evaluation, the direct reduction of iron ore using hydrogen as reduction agent and subsequent steel production in an EAF stands out among all alternative primary production methods.[64]

3.2.3. Hydrogen-Based Steel Production

The main distinction between the BF-BOF and H-DR methods is the substitution of carbon or carbon monoxide as reducing agents by hydrogen to yield water instead of carbon dioxide during the reduction of the iron oxide.[65] The applied hydrogen can be produced by various means, such as nuclear

[56] Cf. Hertwich et al., 2019, p. 15; Milford, Pauliuk, Allwood, & Müller, 2013, p. 3455; Pauliuk & Heeren, 2021, p. 479.
[57] Cf. Pauliuk et al., 2013, p. 3448.
[58] Cf. Xylia et al., 2018, p. 1135.
[59] Cf. Baracchini et al., 2019, p. 79; Demus et al., 2016, p. 569; Fidalgo, Berrueco, & Millan, 2015, p. 279.
[60] Vogl et al., 2021, p. 79.
[61] Cf. Otto et al., 2017; Paltsev, Morris, Kheshgi, & Herzog, 2021; Toktarova et al., 2020.
[62] Cf. Mandova et al., 2019, p. 118.
[63] Cf. Bailera et al., 2021; Fischedick et al., 2014; Weigel et al., 2016.
[64] Cf. Jacobasch et al., 2021, p. 18; Weigel et al., 2016, p. 1074.
[65] Cf. Patisson & Mirgaux, 2020, p. 2.

energy or natural gas with or without CCS technologies.[66] However, the focus lies primarily on the use of green hydrogen, i.e., hydrogen produced by renewable energy sources, as this offers the greatest potential for emission savings. The German government has a similar view: It considers the use of green hydrogen to be the only sustainable option in the long term,[67] which is why it also constitutes the focus of this thesis.

The H-DR process runs as follows: Before actual steel production takes place, the hydrogen must be produced by electrolyzers. These split water (H_2O) into oxygen (O_2) and hydrogen (H_2). At the same time, the iron oxide must be processed into pellets in a pellet plant.[68] Iron production then takes place in a shaft furnace, to which the iron ore pellets are added and subsequently reduced, utilizing the hydrogen as reductant and electricity as energy source.[69] This step is referred to as direct reduction since the iron ore is not melted as in a BF but is solid during the process and hence forms solid iron, called direct reduced iron or sponge iron.[70] The following equation illustrates these processes:[71]

$$Fe_2O_3 + 3H_2 \rightarrow 2Fe + 3H_2O \quad (3)$$

For the final steel production, the produced direct reduced iron must be melted in an EAF. Apart from minor differences in the energy and material flows, this step is similar to secondary steel production. Additionally, steel scrap can be added to reduce the amount of hydrogen required, which affects the quality of the final steel and could thus limit its suitability as a direct substitute for BF-BOF production.[72] In the Figure 4, the major material flows of the described production method are summarized.

The most significant advantage of H-DR production resides in the vast emission savings that can be achieved in primary steel production. Pei et al. (2020) consider the feasible emissions to be around 25 $kgCO_2$-eq/tSteel, less than 2% of BF-BOF emissions, highlighting the enormous potential savings that arise from a production switch.[74] This is only valid if the hydrogen production is entirely green. Assuming hydrogen production with electricity from the German power grid in 2020, this alone would have resulted in indirect emissions of 935 $kgCO_2$ eq/tSteel, fundamentally changing the carbon footprint of this production method.[75] These circumstances highlight a crucial challenge that will have a major impact on the establishment of hydrogen-based steel production and yet remains to be solved: the procurement of the required green hydrogen.

3.2.4. Procurement of Green Hydrogen

In principle, green hydrogen can be procured in two ways. One option is importing hydrogen from countries with large renewable production and export potential. Alternatively, it could be produced domestically.

In the case of domestic hydrogen production, the underlying electricity price turns out to be a decisive factor due to the high electricity consumption of this method. For a price range between € 20-100 per MWh, Vogl et al. (2018) concluded a cost range for H-DR production between € 361-640 per ton of steel.[76] Furthermore, the enormous electricity consumption of the H-DR method is likely to impose an even more significant constraint. In the context of total electricity consumption in Germany, producing the hydrogen required for the H-DR steel would result in enormous burdens to the electricity grid. The production of one ton of steel using the H-DR process consumes roughly 3.5 MWh. Combined with the production volume of nearly 27.98 MMT of primary steel in 2021, the application of H-DR production would amount to a total consumption of nearly 98 TWh,[77] more than 17% of Germany's gross electricity consumption.[78] This picture intensifies considerably if one only takes electricity from renewable energies into account, as would be necessary for the production of entirely green steel and as targeted by the German government: Over 41% of the electricity generated through renewable technologies in Germany in 2021 would be required to power primary steel production in its current volume.[79]

Such problems are not expected for imported hydrogen since it can be assumed that it is supplied by regions with enormous production potential for green hydrogen. Such assumptions have already been made by the German government as well. In the National Hydrogen Strategy, it concluded that "(...) the domestic generation of green hydrogen will not be sufficient to cover all new demand, which is why most of the hydrogen needed will have to be imported."[80], thus raising the necessity for corresponding supply infrastructure.

The establishment of such infrastructure is currently subject to extensive interest in literature and is still associated with many uncertainties. These relate, for example, to the source of supply, the form of transport, and its temporal availability. Brändle et al. (2020) investigated the global hydrogen export potential of different countries in terms of volume and costs by using newly built or refurbished pipelines. The

[66] Cf. Germeshuizen & Blom, 2013, p. 10671; Toktarova, Göransson, & Johnsson, 2021, pp. 2-3.
[67] Cf. BMWI, 2020b, p. 2.
[68] Cf. Pei, Petäjäniemi, Regnell, & Wijk, 2020, p. 9; Toktarova et al., 2021, p. 3.
[69] Cf. Vogl et al., 2018, pp. 737-738.
[70] Cf. Patisson & Mirgaux, 2020, p. 2.
[71] Otto et al., 2017, p. 10.
[72] Cf. Kirschen et al., 2011, p. 6151; Vogl et al., 2018, pp. 739, 743.
[73] Own illustration based on process description above.
[74] Cf. Pei et al., 2020, p. 7.
[75] Electricity consumption: 50.1 kWh/kgH_2, H_2 consumption: 51 kg/tSteel, grid emission factor: 0.366 $kgCO_2$-eq/kWh. Cf. Brändle, Schönfisch, & Schulte, 2020; UBA, 2021, p. 9; Vogl et al., 2018, p. 739.
[76] Cf. Vogl et al., 2018, p. 744.
[77] Cf. Pei et al., 2020, p. 8; Vogl et al., 2018, p. 739; WV Stahl, 2022, p. 1.
[78] Gross electricity consumption in Germany, 2021: 565.3 TWh, cf. AGEB, 2021.
[79] Share of renewable electricity: 41.9%, cf. AGEB, 2021.
[80] BMWI, 2020a, p. 2.

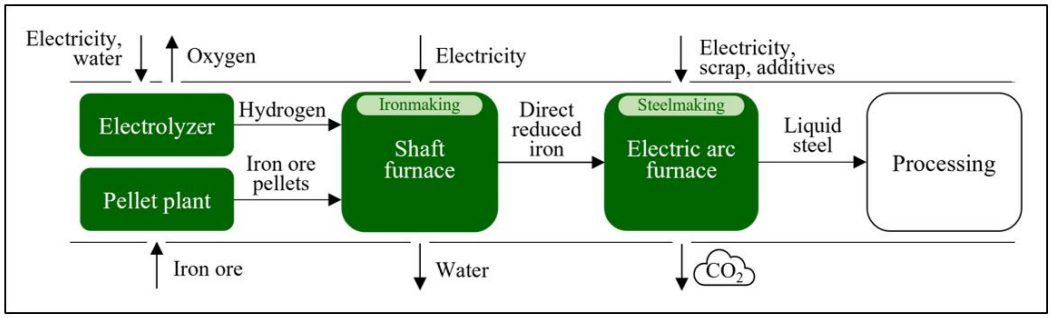


Figure 4: Primary steel production within the H-DR production route.[73]

researchers deduced that onshore wind sources from northwestern Europe or photovoltaic sources from southern Europe are well suited for exports of green hydrogen to Germany. If transport costs are low, Morocco or Algeria offer great potential, too.[81] The German federal government intends to foster such infrastructure as well as the establishment of an international market for hydrogen in the National Hydrogen Strategy, and one year after its publication reported about initial initiatives with the purpose of importing hydrogen from non-European countries.[82]

Nevertheless, concrete developments are not apparent. When, from where, and in what volume green hydrogen could be imported to Germany is therefore still subject to great uncertainty. Similarly, the costs of such imports are not yet precisely foreseeable, even though studies have identified these as a driver that will significantly determine the future cost competitiveness of H-DR steel.[83]

An initial concept for import infrastructure for hydrogen is provided by the European Hydrogen Backbone Initiative, which was jointly founded by several European gas infrastructure companies. These envision to successively expand a European pipeline network and link the first industrial regions, including parts of Germany, as early as 2030.[84]

Consequently, it can be concluded that the procurement and cost of hydrogen will play a major role in shaping the future role of H-DR production. Hydrogen procurement is subject to various limitations in this regard: While importing holds out the prospect of a fully green and low-cost option in sufficient quantities, it requires appropriate infrastructure. Domestic production may be available at an earlier stage but is expected to be more expensive and can only provide limited quantities. Thus, potential could arise from domestic production as a transitional technology until import infrastructure is available.

3.3. Internal Dynamics

This section analyzes the foreseeable developments for implementing hydrogen-based production within the steel industry. Especially the willingness of German steel producers and steel consumers to transform as well as developments thereof will be considered. The analysis of steel producers relates mainly to the three largest producers in Germany: thyssenkrupp, ArcelorMittal, and Salzgitter AG. As already outlined in Chapter 3.1, these account for the majority of total German steel production.

Fundamentally, each of the three producers has formulated the goal of entirely carbon-neutral steel production, which is equivalent to a commitment to depart from conventional BF-BOF production due to its high emission intensity. ArcelorMittal plans to become carbon-neutral by 2050, and thyssenkrupp intends to produce only climate-neutral steel by 2045.[85] Salzgitter even opts to transform its entire primary production until 2033, saving 1% of Germany's total emissions solely in the course of that. Besides introducing H-DR production, Salzgitter plans to substantially increase its secondary steel production.[86]

Considering current H-DR projects on a general level, promising pilot projects in Sweden are particularly noteworthy. There, the decarbonization of the steel industry is receiving attention, as the domestic steel industry is responsible for 10% of all emissions and thus plays an essential role in achieving the target of climate neutrality by 2045.[87] Karakaya et al. (2018) concluded that Swedish companies, governmental as well as research institutions "(...) strongly collaborate to drive the transition towards hydrogen-based direct reduction technology."[88] In particular, the HYBRIT project, a joint venture between Swedish companies SSAB, LKAB, and Vattenfall, is receiving broad attention. This project plans to bring green steel to market through demonstration plants starting in 2026 and to establish industrial production between 2030 and 2040 to fully produce carbon-neutral steel in 2045.[89] The H_2 Green Steel project, also based in Sweden, is even more ambitious and plans to start green steel production in 2024, aiming at producing five million tonnes per year by 2030[90] - more than the total Swedish

[81] Cf. Brändle et al., 2020, pp. 24-25.
[82] Cf. BMWI, 2021b, pp. 1-2.
[83] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 167; Mayer, Bachner, & Steininger, 2019, p. 1520; Vogl et al., 2018, p. 744.
[84] Cf. Wang, van der Leun, Peters, & Buseman, 2020, p. 4.
[85] Cf. ArcelorMittal, 2021, p. 31; Thyssenkrupp, 2022.
[86] Cf. Salzgitter AG, 2022b.
[87] Cf. Pei et al., 2020, p. 2.
[88] Karakaya et al., 2018, p. 662.
[89] Cf. Pei et al., 2020, p. 10.
[90] Cf. H2GS, 2022.

production volume to date.[91]

Although German steel production is more than eight times larger than its Swedish counterpart,[92] comparatively small-scale projects exist. The following table provides an overview of the mentioned Swedish projects and the largest announcements for hydrogen-based iron or steel production in Germany up to date:

The observed projects, as well as the formulated emission targets, indicate a willingness of steel producers to transform their current production processes by implementing H-DR technology. Salzgitter AG and thyssenkrupp, in particular, are targeting significant production volumes within the next decade. Nevertheless, contrary to the Swedish projects, concrete commitments only account for small parts of the total production volume.

Furthermore, the role of steel consumers constitutes an essential factor. Representatives of the steel industry argue "(...) that substantial additional costs could not be borne by steel producers, as they operate on increasingly miniscule margins in a globally competitive market."[94] Because hydrogen-based production of green steel is initially most likely associated with higher costs, the willingness of steel consumers to purchase the product constitutes an essential factor for its success. In principle, German steel producers compete with international producers, as customers can freely choose between the alternatives. The decline in German steel production in recent years despite a growing global production indicates that consumers have increasingly chosen cheaper alternatives from abroad and thus preferred lower-cost options.[95]

Nonetheless, initial positive signals can already be observed. First analyses of primary steel-consuming industries point to favorable framework conditions. Rootzén and Johnsson (2016) reported that the additional cost of using green steel in the automotive industry would result in an increase of less than 0.5% in the cost of a mid-size car and thus would only moderately influence purchasing decisions.[96] This example is reinforced by Muslemani et al. (2021), who identified potential for an increased willingness to pay for green steel most likely to develop in the automotive sector.[97] Furthermore, initial procurement commitments exist from various companies that attach additional value to green steel compared to its conventional equivalent. Examples are listed in the following table:

However, these commitments are still within a somewhat limited scope, from which a comprehensive demand cannot be derived. The extent to which the added value through the green property of the steel can counteract higher associated costs on a large scale cannot be precisely determined at present. In studies of comparable thematic areas, positive indications are found that could favor such development. Exemplary, these studies have derived that consumers show an increased willingness to pay for green electricity.[99] If such effects were to be transferred to the steel market, significant opportunities for accelerating the transformation could arise.

Steel producers likewise consider the development of dedicated markets for green steel as a major opportunity. On this occasion, the German Steel Federation calls for suitable policy instruments such as quotas or setting incentives to stimulate demand.[100] The role of political support is analyzed in more detail in the next section, in which a perspective on the current policy frameworks and foreseeable developments thereof is provided.

3.4. Political Environment

Political directives to which the German steel industry is subject exist at the European and national levels. In the context of this thesis, climate policy aspects will be discussed in particular, as these are expected to significantly contribute to the development of hydrogen-based steelmaking in the coming years.

3.4.1. European Policy

In 2019, the European Commission presented the European Green Deal, a concept that envisions making the European Union climate-neutral by the year 2050. This concept also defines targets for the steel industry. It emphasizes the decarbonization of the steel industry as an essential part of total decarbonization and that new emission-free technologies are to be promoted for introduction starting in 2030. Hydrogen-based steel production is mentioned as a potential technology for such decarbonization.[101] Factors such as these set the European Union apart from other regions: Based on an analysis of the current framework conditions, Arens et al. (2021) concluded that the European Union currently offers steel producers the globally most promising environment for a transformation of steel production in the coming decades.[102]

Part of the current regulatory framework of the European Union is its Emission Trading System (EU ETS), to which steel producers are subject. Within this system, they are in principle obliged to cover their emissions by purchasing allowances. The actual exposure to this regulation is severely limited at the moment, as will be explained in the following.

Being a trade-intensive industry, European steelmaking directly competes with international competitors. For this reason, policymakers fear that the EU ETS would increase

[91] Swedish steel production in 2019: 4.7 MMT, cf. Worldsteel, 2009-2022, p. 1.
[92] Cf. Worldsteel, 2009-2022, p. 1.
[93] Cf. ArcelorMittal, 2021, p. 43; H2GS, 2022; HYBRIT, 2021; Salzgitter AG, 2022b; Thyssenkrupp, 2021, p. 67.
[94] Muslemani et al., 2021, p. 9.
[95] As outlined in Chapter 3.1.
[96] Cf. Rootzén & Johnsson, 2016, p. 1.
[97] Cf. Muslemani et al., 2021, p. 9.
[98] Cf. Faurecia, 2021; Mercedes-Benz Group, 2021; Miele, 2021; Salzgitter AG, 2022b; Salzgitter AG, 2022a; Scania, 2021; Schaeffler, 2021; SSAB, 2022; Volvo Group, 2022.
[99] Cf. Sundt & Rehdanz, 2014, p. 16.
[100] Cf. WV Stahl, 2021c, p. 3.
[101] Cf. EC, 2019b, pp. 1-8; EC, 2020, p. 1.
[102] Cf. Arens et al., 2021, p. 8.

Table 1: Announced projects for hydrogen-based steelmaking.[93]

Country	Company	Project name	Year online	Iron/steel volume (MMT/year)
Sweden	H_2 Green Steel	H_2GS	2024	5 (by 2030)
Sweden	SSAB	HYBRIT	2026	2.7 (by 2030)
Germany	ArcelorMittal	Hamburg H_2	2023-2025	0.1 (by 2023-2025)
Germany	Salzgitter AG	SALCOS	2025	>3 (by 2033)
Germany	thyssenkrupp	tkH_2Steel	2025	3 (by 2030)

Table 2: Public commitments to purchase green steel.[98]

Company	Supplier	Year	Volume (MMT/year)
Mercedes-Benz	H_2GS	2025	N/A
Scania	H_2GS	N/A	N/A
Schaeffler	H_2GS	2025	0.1
Faurecia	SSAB	2026	N/A
Polestar	SSAB	N/A	N/A
Volvo	SSAB/ Ovako	2022	N/A
BMW	Salzgitter AG	2026	>0.5
Miele	Salzgitter AG	2021	>0.288 (low carbon instead of entirely green steel)
Volkswagen	Salzgitter AG	2025	N/A

the risk of carbon leakage. In this case, this threat refers to the relocation of steel production capacities abroad due to lower production costs, which would result in a loss of the industry from the perspective of the European Union. At the same time, emissions would still be generated elsewhere, thus counteracting climate policy efforts.[103] Therefore, the EU ETS provides industries exposed to the risk of carbon leakage with partially or entirely free emission allowances to prevent them from being disadvantaged in international competition. Steel production has been classified as such industry by the European Commission.[104] Under the provisions of the currently active Phase 4 of the EU ETS, which runs from 2021 to 2030, it is planned that steel producers will thus receive free allowances for all generated emissions.[105] Therefore, no direct cost pressure for steel producers incurred by the EU ETS regime in the next few years is apparent at the current time.

Additionally, issues related to the distribution of emissions can be observed. Although the allowances granted to steel producers are measured based on a benchmark set by the most emission-efficient producers, European steel producers have consistently received excessive free allowances since the EU ETS was launched: In 2019, the free distributions covered around 27% more emissions than were verified for steel producers in the European Union, resulting in windfall profits, which they obtain by selling the allowances.[106] Presumably, this can be attributed to lobbying activities by steel producers, through which they strategically exaggerated their vulnerability to the EU ETS in the past and thus successfully exerted influence on its design.[107]

Considering the roll-out of H-DR technology, the cost of allowances to compensate for emissions is a factor that could accelerate its competitiveness towards BF-BOF production and thus contribute significantly to its success.[108] However, to generate such an effect, the EU ETS in its current form turns out to be insufficient. Various studies argue that the EU ETS has not yet resulted in adequately high CO_2 prices to incentivize the application of more expensive low-carbon alternatives to substitute basic materials or energy-intensive technologies in general.[109] Furthermore, the high volatility of CO_2 prices covered by the EU ETS shapes the investment base of capital-intensive projects as uncertain and poorly suited for making the necessary, far-reaching investment decisions.[110]

Since the current EU ETS regulation can be assessed as rather inefficient concerning the uptake of H-DR production, alternative policy instruments have already been suggested to foster H-DR technology more effectively while still preventing carbon leakage. The European Commission presented a concrete proposal in July 2021 as part of its Fit-for-55 package.[111] The package proposes a phase-out of

[103] Cf. Branger, Quirion, & Chevallier, 2016, pp. 109-110.
[104] Cf. EC, 2019a, p. 25.
[105] Cf. EC, 2021d, p. 221.
[106] Cf. Carbon Market Watch, 2016, p. 3; EEA, 2021.
[107] Cf. Okereke & McDaniels, 2012, p. 9.
[108] Cf. Jacobasch et al., 2021, p. 16; Vogl et al., 2018, p. 744.
[109] Cf. Sartor & Bataille, 2019, p. 5; Vogl et al., 2021, p. 81.
[110] Cf. Sartor & Bataille, 2019, p. 6; Vogl et al., 2021, p. 81.
[111] Cf. EC, 2021a; EC, 2021c, p. 12.

all free allowances between 2026 and 2035 and instead to establish a Carbon Border Adjustment Mechanism for various products. In the case of steel, such a mechanism would impose tariffs on steel imported into Europe to offset the additional allowance costs of European producers. Conversely, when European producers export their steel, they would be reimbursed for the allowance costs to ensure competitiveness in the international market. Even if this alternative seems effective in fully internalizing CO_2 costs, its introduction would be associated with considerable administrative effort: For each product concerned, extensive knowledge about its emission intensity would have to be available, and trade law disputes would be likely.[112] As this model only constitutes a proposal at the present time, a continuation of the free allowances regime for steel producers seems to be the most likely option at the European level.

3.4.2. German Policy

German policy has set climate targets that exceed European regulations. In 2021, the Federal Climate Change Act was amended, tightening the national goals. There, Germany has set the targets of reducing 65% of 1990 emissions by 2030, becoming climate-neutral by 2045, and achieving negative greenhouse gas emissions from 2050 onwards.[113] Specific emission targets were also set for all major sectors. As steel production was responsible for around 28% of all industrial emissions in 2020,[114] regulations for the industry sector are particularly relevant. Between 2010 and 2019, the industrial sector reduced its emissions by less than 3%, from 188 to 183 MMT CO_2-eq, thus contributing minimally to Germany's emissions reductions to date.[115] However, considering the update of the Federal Climate Change Act, it becomes evident that much higher emission reductions are anticipated until 2030. By then, total industrial emissions are to be reduced by more than 35% from 2019 levels to 118 MMT CO_2-eq,[116] from which a significant emission reduction pressure on German steel producers can be derived.

The significant role the German steel industry will play in the decarbonization of the overall economy has already been recognized by the German government in 2020 with a concept developed jointly with steel producers, the Steel Action Concept. In this, as well as in the National Hydrogen Strategy, the H-DR production method is considered the most promising decarbonization alternative.[117] In the Steel Action Concept, the German government signals a strong willingness to foster hydrogen-based steel production and already envisions support within the framework of other policy instruments, such as the promotion of markets for green steel, the establishment of adequate energy infrastructure and markets for hydrogen, among others. Although the Steel Action Concept includes the intention to continue the free allowances of the EU ETS in its current form, it additionally indicates an openness towards other carbon leakage prevention instruments without naming any definite plans.[118]

More specific information was published one year after the publication of the Steel Action Concept: In 2021, the Federal Ministry for Economic Affairs and Energy listed electricity price compensations, a reduction in the levy to support renewable energies, and the free allocation of allowances within the EU ETS as measures already active to support steel producers. Furthermore, an announced support package intended for the entire industrial sector could initially affect the German steel industry. From 2022 to 2024, the German government announced funding of five billion euros for promoting the application of hydrogen or to test the suitability of carbon contracts for difference (CCfD) to initiate the transformation as part of pilot projects, among others.[119]

The concept of CCfDs is regarded as an efficient instrument for the large-scale commercialization of promising industrial technologies.[120] In the given context, such a contract could consist of an agreement between the regulator, such as the German government, and steel producers to subsidize H-DR projects. For this occasion, a strike price, which is the CO_2 price the H-DR plant needs to become competitive with conventional steel production, and a period in which this strike price is guaranteed, are first agreed upon. If the actual CO_2 price is below the strike price within this period, the steel producer receives payments from the regulator for each avoided quantity of emissions. If the CO_2 price is above the strike price, the producer conversely has to make payments to the regulator.[121] To ensure the efficient formation of the strike price, CCfDs could be allocated among producers through tendering processes.[122] This policy measure entails several advantages. In addition to offsetting increased operating costs, the precisely defined conditions reduce uncertainty for producers regarding the development of CO_2 prices as well as future policy developments, resulting in better investment conditions. Furthermore, basing the payment on avoided emissions creates incentives for the project to be implemented successfully. From a regulator's perspective, opportunities arise to recoup expenditures if CO_2 prices rise above the strike price, limiting the threat of over-subsidization. A potential weakness may be found in the complexity of the design, making this instrument most suitable for large-scale projects.[123]

In conclusion, it can be noted that while ambitious emission reductions have been formulated for the industrial sector by German policymakers, a high degree of willingness to support the transformation of steel production is also evident. A combination of several instruments seems most likely

[112] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 106.
[113] Cf. Federal Climate Change Act, Section 3.
[114] Cf. DEHst, 2014-2021, p. 28; UBA, 2022.
[115] Cf. UBA, 2022.
[116] Cf. Federal Climate Change Act, Annex 2.
[117] Cf. BMWI, 2020b, pp. 2-7; BMWI, 2020a, p. 2.
[118] Cf. BMWI, 2020b, pp. 13-17.
[119] Cf. BMWI, 2021a, p. 2.
[120] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 110.
[121] Cf. Vogl et al., 2021, p. 84.
[122] Cf. Sartor & Bataille, 2019, p. 10.
[123] Cf. Agora Energiewende und Wuppertal Institut, 2019, pp. 110-113; Richstein, 2017, p. 16.

at present, but the use of CCfDs, the application of which has already been announced in pilot projects, should be particularly emphasized. As current measures are still far from a comprehensive commercial rollout, and concrete projects for supporting the steel industry have not yet been published, this perspective is still subject to increased uncertainty.

3.5. Summary of the Initial Situation and Identification of Major Drivers

After analyzing the current and foreseeable framework for the introduction of H-DR production from various perspectives, the key findings are summarized in this section. Firstly, the aim is to obtain initial qualitative findings on the first research question and thus to identify factors and interrelationships that are most likely to influence the diffusion of hydrogen-based steel production. Furthermore, these findings will serve as the foundation for quantitative modeling in the next step of this thesis.

A key result that emerges from the conducted multi-perspective analysis lies in the conclusion that the emission targets set by German policymakers are unlikely to be achieved with the currently applied BF-BOF production method without reducing production volume. To realize a sustainable alignment of steel production with the overall German decarbonization pathway, an urgent need to transform primary steel production can thus be derived. Although various alternatives exist for substituting the BF-BOF method, hydrogen-based direct reduction promises great potential primarily due to its exceptionally low emission intensity. Additionally, H-DR production emerges as the currently most favored technology, as it forms a pivotal role in key political concepts as well as in the corporate strategies of the largest German steel producers. Thus, it can be concluded that extensive development of the H-DR method is likely to take place, standing out from alternatives in all observed aspects.

However, one factor that could influence the scale of H-DR production is found in the future role of secondary steel production. Various studies hold out the prospect of a global expansion of the latter in the coming decades. Furthermore, first steel producers expect an expansion of their secondary production to achieve the set emission targets. As such, the prospect of an increasingly dominant role in EAF production within the German steel industry represents a plausible option.

Regarding H-DR production, several major uncertainties exist, which could significantly impact the success of this technology. Above all, the expected higher production costs result in substantial disadvantages. The German steel industry currently finds itself in a disadvantageous position to manage this burden. Whereas the steel industry, in general, provides rather unattractive investment conditions, German steel producers have been additionally exposed to tough international competition in recent years, resulting in declining production volumes and sales. Achieving competitive production costs between H-DR and BF-BOF thus represents an essential precondition for adopting the new technology and preventing the migration of steel producers due to increasing cost pressures. Two factors were identified as having the most substantial impact on future production costs: On the one hand, the costs for the required hydrogen, whose future reduction could make the H-DR more attractive, and on the other hand, costs for the compensation of generated emissions through CO_2 prices, which would primarily result in increased costs of BF-BOF production.

Regarding the procurement of hydrogen, it appears that the import of hydrogen will probably be superior to domestic production in Germany solely due to the capacity limitations of the German energy grid. However, such large-scale imports are not yet foreseeable and are subject to significant uncertainty, which may allow domestic production to act as a transitional technology in the short term.

Besides the actual development of CO_2 prices in the EU ETS market, an additional factor is found in their applicability to steel production through political regulation. The analysis of the current policy framework showed that the protection of steel production from carbon leakage has so far played a pivotal role in policy measures. The resulting distribution of free allowances does not indicate any additional cost pressure for BF-BOF production until at least 2030 and results in a lack of stimulation of the EU ETS for steel production. Nevertheless, the European Commission has already submitted alternative proposals for an early phase-out of free allowances and a simultaneous introduction of a Carbon Border Adjustment Mechanism that might play a role in the future. German policymakers are also expressing great willingness to support H-DR production. Extensive support measures such as the creation of required energy infrastructure, the promotion of green lead markets, and, above all, the establishment of CCfDs are suggested. Still, specific projects have yet to materialize. Accordingly, the future design of the political framework in terms of its type and scope is a factor that is expected to impact H-DR diffusion heavily.

Lastly, the market potential of green steel remains an open question which might lead to an acceleration of H-DR diffusion. Several companies have already made initial commitments to purchase green steel. However, it is to be clarified to what extent this will trigger a comprehensive increase in the willingness to pay on the part of consumers, leading to the establishment of separate markets that soften direct competition with conventional steel.

In summary, it can be concluded that a complex picture of the framework conditions for H-DR diffusion emerges. Various factors that are often subject to considerable uncertainty have been identified as drivers exerting a decisive influence on the future role of hydrogen-based steel making. Table 3 provides an overview of the discussed factors.

Table 3: Identified major drivers for H-DR diffusion

Dimension	Driver
Technological	Expansion of secondary production
Industry-internal	Market potential of green steel
Costs	Development of Hydrogen costs Domestic hydrogen production feasibility Import infrastructure availability Development of CO_2 prices
Political	Development of the free allowances regime Overall type and scope of policy measures

4. Scenario Development

After the prevailing framework conditions for the diffusion of H-DR steelmaking were defined in the first part of this thesis, these will now be incorporated into a quantitative model. Based on this model, the aim is to develop realistic scenarios of H-DR diffusion in Germany to identify critical relationships and gain insights for policymaking, as pointed out in the research questions. Figure 5 summarizes the developed model, followed by a description of the individual steps and the input variables employed.

4.1. Methodology

Fundamentally, the model considers the period from 2022 to 2050 and links directly to historical values underlying until 2020 or 2021, depending on availability. The limitation to 2050 is implemented since many input variables, such as the selected CO_2 prices and hydrogen costs, are only available up to this year.[125] Furthermore, the year 2050 is often the boundary for political emission and climate targets. For instance, Germany's Climate Change Act sets concrete targets up to 2045, while for the period after 2050, the only statement is that negative greenhouse gas emissions are to be realized.[126]

The model consists of three successive steps, each achieved by incorporating additional variables. The foundation is formed by a tipping point analysis, in which forecasts for the price development of the various production methods are compiled by including pivotal cost drivers. In addition to the established technologies, the costs for hydrogen-based steel production are taken into account, each for importing and domestically producing the hydrogen. Based on these cost projections, a general diffusion scenario of H-DR production as replacement for BF-BOF is then simulated. The final step provides an approach to model the impact of strict compliance with the defined annual emission budgets on actual production volumes and the loss of primary steel production.

Each scenario is subject to a policy framework, which serves as the lead indicator since it is expected to impact other input variables and the overall development most decisively. Quantitatively, the policy framework is reflected by various options for the future distribution of free allowances. Three different variations are analyzed: a continuation as currently anticipated, an early phase-out, and the complete absence of free allowances. Within the first option, it is assumed that steel producers can compensate all emissions at no cost until 2030, followed by a linear phase out until 2040. The ten-year duration of this process is derived from the European Commission's early phase-out proposal contained in the Fit-for-55 package, which estimates this period for the phase-out. This proposal further describes the second alternative considered. For the H-DR diffusion, this would entail an earlier onset of support, as it foresees a phase-out of free allowances between 2026 and 2035 and would likely be accompanied by the introduction of new policy instruments.[127] The final policy framework is the complete absence of free allowances for steel producers. This alternative is a somewhat unrealistic assumption, as such developments are not apparent at the current time. However, this variant will serve as a benchmark as it allows drawing interesting conclusions, especially on the effects of the distribution of free allowances.

4.1.1. Tipping Point Analysis

The tipping point analysis aims to model the production costs of the observed production routes per ton of steel as an aggregate of the most critical cost factors. For each scenario, the goal is to identify the tipping point - the year in which cost equality between H-DR and BF-BOF production is achieved. The analysis is undertaken in real euros with the base year 2020. Currency transformations are conducted based on historical exchange rates of the European Central Bank, and the adjustment of cost data to the underlying base year is carried out using the wholesale price index by the Federal Statistical Office.[128] A detailed list of the underlying material quantities and price inputs is provided in Appendix 1 to Appendix 4.

[124] Own illustration.
[125] Cf. Brändle et al., 2020, p. 1; IEA, 2021, p. 329.
[126] Cf. Federal Climate Change Act, Section 3.
[127] Cf. EC, 2021a.
[128] Cf. Destatis, 2022, p. 7; ECB, 2022.

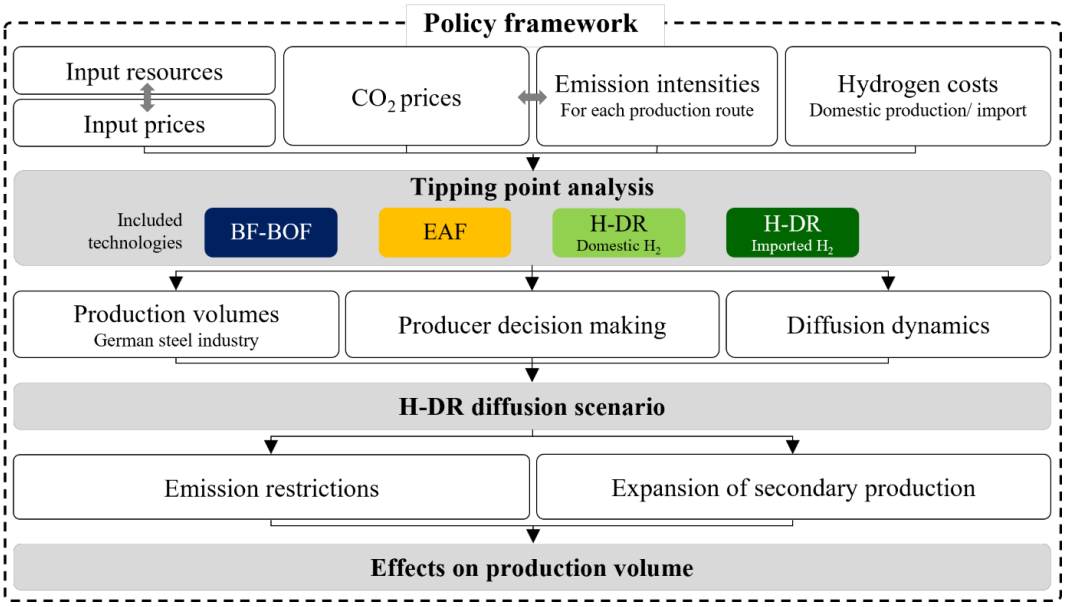


Figure 5: Methodology for modeling the diffusion of H-DR steelmaking in Germany.[124]

General Input Resources

The consumption of materials constitutes the largest part of the production costs of all methods. The specific material flows were identified through literature research as well as through comparison with assumptions of similar papers. In order to derive the material costs, these quantities are linked to cost data. Due to exceptionally high fluctuations in various raw material costs, particularly since the beginning of the Covid-19 pandemic, the prices underlying the model were extrapolated linearly over the entire observed period based on the trend of historical market prices between 2010 and 2021. A cost progression corresponding to the average of available cost data was assumed for cost factors for which extensive historical data are not available.[129]

Besides the material costs, additional cost components are taken into account. For the calculation of the capital expenditures, data of Woertler et al. (2013) is applied, as it is also underlying in the models of Fischedick et al. (2014), Toktarova et al. (2020), and Vogl et al. (2018).[130] Assumptions following the methodology of Vogl et al. (2018) are employed to determine annual capital costs: The authors assigned a lifetime of 20 years to all production facilities and applied an interest rate of 5%.[131] Furthermore, operations and maintenance costs are accounted for at 3% of capital expenditures, as introduced by Fischedick et al. (2014).[132] As indication of labor costs, the analysis of Medarac et al.(2020) is referenced. Due to lack of data availability, it is assumed that the labor costs of H-DR and BF-BOF production are equal.[133] Since no comprehensive historical data are available for steel production-specific capital and labor costs, these are assumed to remain constant.

Because the underlying CO_2 prices within the EU ETS and the costs for hydrogen have been identified as major drivers for the future H-DR development, these are considered in a more detailed analysis.

CO_2 Prices

CO_2 prices are expected to rise, thus having the primary effect of increasing the costs of the most emissions-intensive BF-BOF production. Since none of the observed production methods is consistently carbon-neutral, CO_2 prices will impact all of them accordingly, albeit less than for BF-BOF production. For specific input values of CO_2 prices, forecasts of the International Energy Agency (IEA) within the framework of its World Energy Model are adopted. In this model, the IEA forecasts CO_2 prices in different scenarios for 2030, 2040, and 2050. By linearly interpolating the years in between, these projections are included in the model as illustrated in Figure 6.

For the Net Zero Emissions by 2050 (NZE) scenario, the IEA's modeling is subject to a pathway in which the global energy sector will be carbon-neutral by 2050. In the Announced Pledges (APS) scenario, the IEA assumes that current climate commitments made by all governments worldwide will be implemented as announced. The Stated Policies (STEPS) scenario "(...) reflects current policy settings based

[129] This concerns oxygen and the steel production-specific materials of fluxes and graphite electrodes.

[130] Cf. Fischedick et al., 2014, pp. 577-578; Toktarova et al., 2020, p. 16; Vogl et al., 2018, p. 741; Woertler et al., 2013, p. 22.

[131] Cf. Vogl et al., 2018, p. 739.

[132] Cf. Fischedick et al., 2014, p. 577.

[133] Cf. Medarac et al., 2020, pp. 10-12.

[134] Cf. ICAP, 2022; IEA, 2021, p. 329; own calculations. 2021 value based on interpolation, as the actual development was exceptionally volatile. All numerical values are listed in Appendix 4.

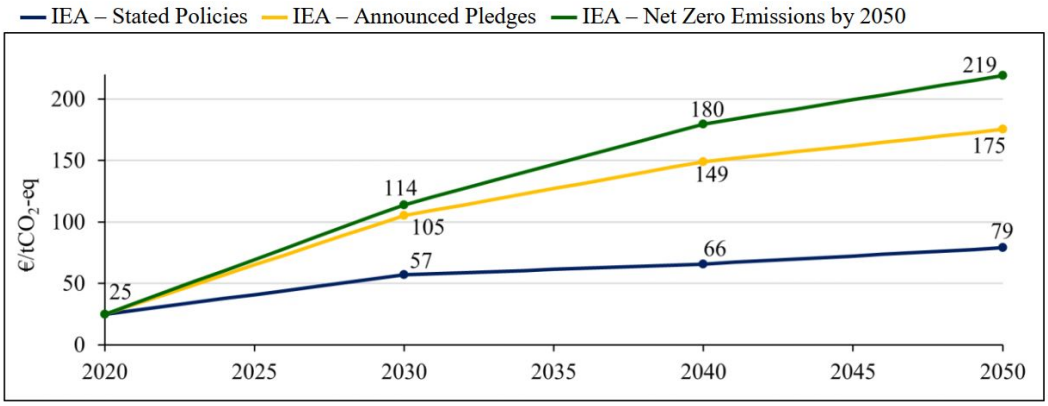


Figure 6: Underlying forecasts of CO_2 prices.[134]

on a sector-by-sector assessment of the specific policies that are in place, as well as those that have been announced by governments around the world."[135]

Hydrogen Costs

Contrary to CO_2 prices, Hydrogen costs are expected to decline and thus increase the attractiveness of H-DR production. Factors such as economies of scale, supply infrastructures, and technological advances are anticipated to reduce the costs of green hydrogen, decreasing the costs of green steel alongside. The work of Brändle et al. (2020) is employed to model hydrogen costs. Within their study, the authors provide a detailed long-term and scenario-based forecast for the supply costs of hydrogen in more than 80 countries. Numerous techno-economic assumptions include learning curves, transport distances, and efficiency improvements, as well as capital requirements of electrolyzers.[136] The values underlying are obtained from the Tool for Costs of Hydrogen, co-published by the authors alongside the paper.[137] To preserve the import costs for sourcing in Germany, the average price of the ten most competitive hydrogen sources is applied up to 2050. Based on the data on low-cost pipeline transport of hydrogen, one data series is extracted each for optimistic and baseline assumptions.[138] Analogously, the average costs of all production options, without the influence of different transport alternatives, are considered as domestic production costs in Germany for each year until 2050 for optimistic and baseline assumptions. The optimistic data series include improved techno-economic assumptions such as underlying learning curves or electrolyzer capital costs.[139]

For the imported hydrogen, it is assumed that it was produced exclusively with renewable energies since it originates from regions with considerable production potential and offers the prospect of sufficient availability. However, the production of hydrogen in Germany is subject to the assumption that it cannot be accomplished exclusively with renewable capacities due to limitations of the German electricity grid, as already discussed in Chapter 3.2.4. For this reason, it is assumed that hydrogen can only be produced domestically with electricity from the national power grid, which results in the following modification for the domestic production costs: In addition to the hydrogen costs derived from Brändle et al. (2020), compensation costs for the indirect emissions caused by consuming electricity from the German power grid will be considered. This calculation is based on the projections of electrolyzer efficiency improvements by the IEA (2019), which is also applied by Brändle et al. (2020), and a projection of the emission intensity of the German electricity grid.[140] The German grid emission factor is assumed to linearly decline from 366 gCO_2-eq/kWh in 2020 to zero in 2045 in line with the net neutrality target in the Federal Climate Change Act, as similarly modeled, for example, by Fischedick et al. (2014).[141] The applied logic yields the hydrogen costs shown in Figure 7.

Emission Intensities

The final variable for the tipping point analysis is the emission intensity of the individual production methods. Combined with the underlying CO_2 prices, this factor directly impacts the costs of the produced steel, as long emissions are not fully covered by free allowances. Different developments are foreseeable for the individual emission intensities, as explained in the following.

Only limited potential for future reductions in its emission intensity can be identified in BF-BOF production. An indicator for such improvements is provided by the benchmarks set

[135] IEA, 2021, p. 27.
[136] Cf. Brändle et al., 2020, pp. 7-16.
[137] The third version of the tool from March 2021 was used.
[138] The model includes the transport alternatives of retrofitted pipelines, low-cost, and high-cost pipelines, of which the low-cost pipeline was selected as intermediate option.
[139] Cf. Brändle et al., 2020, pp. 10-12.
[140] Cf. IEA, 2019, p. 44.
[141] Cf. Federal Climate Change Act, Section 3; Fischedick et al., 2014, p. 572; UBA, 2021, p. 9.
[142] Cf. Brändle et al., 2020; own calculations. For reasons of clarity, only the baseline values are visualized for domestic production. All numerical values are listed in Appendix 4.

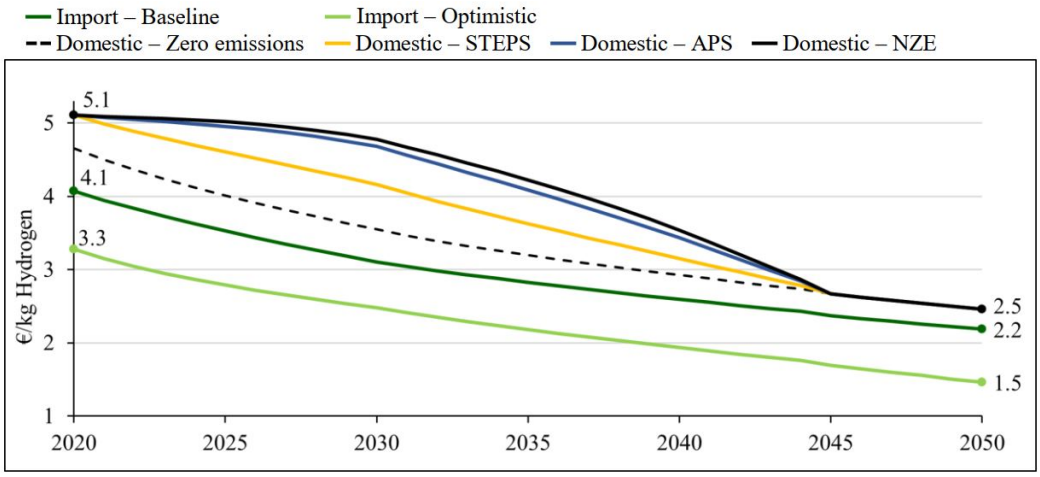


Figure 7: Underlying forecasts of hydrogen costs.[142]

by the EU ETS framework. Within the current Phase 4 of the EU ETS, an annual reduction of BF-BOF emissions by 0.2% is expected due to technological improvements.[143] This model adopts this assumption to forecast BF-BOF emissions over the observed period. Hence, a reduction of BF-BOF emissions from around 1,800 $kgCO_2$-eq/tSteel in 2020 to almost 1,700 $kgCO_2$-eq/tSteel in 2050 is underlying.

The emissions from secondary steel production in the EAF first need to be divided into an indirect and a direct component. The indirect emissions result from consuming electricity supplied by the German power grid and are thus reduced alongside the grid emission intensity from 244 $kgCO_2$-eq/tSteel in 2020 to zero in 2045.[144] The direct emissions result from adding fossil fuels during the melting process in the EAF. In 2020, direct emissions are assumed to be at 100 kg/tSteel.[145] However, as different studies suggest replacing fossil fuels with sustainable alternatives, future zero-emission steel production in the EAF is a reasonable prospect.[146] Therefore, it is assumed that direct emissions, alongside indirect emissions, will be reduced linearly to zero by 2045.

Since the H-DR production is also subject to a melting process in an EAF, it initially consists of the same emission dynamics as secondary production. Additionally, the H-DR method requires more electricity because the shaft furnace must be operated. Secondly, the EAF within the H-DR production consumes slightly more electricity.[147] The emission intensity of the H-DR production with imported hydrogen, which is entirely produced from renewable energy, is thus only about 149 $kgCO_2$-eq/tSteel higher than the emissions of secondary production in 2020. Significantly different emissions persist if the hydrogen is produced domestically with electricity from the German power grid, as assumed. Due to the high electricity consumption of the electrolyzers, additional indirect emissions of 935 $kgCO_2$-eq/tSteel arise at the emission intensity of the German electricity grid in 2020, worsening the climate balance of this alternative.[148] Nevertheless, this method features vast improvements under the assumptions adopted. By decarbonizing the power grid and increasing the efficiency of electrolyzers, this alternative also allows carbon-neutral steel production by 2045. Thus, the emission intensity projections can be summarized as follows in Figure 8.

4.1.2. Diffusion Scenario of Hydrogen-Based Steelmaking

As shown in the model methodology in Figure 5, the tipping point analysis serves as the foundation for the H-DR diffusion scenarios. In these scenarios, the H-DR roll-out is simulated according to plausible diffusion dynamics. A description of the underlying inputs is provided in the following.

Producer Decision Making

Even though the market potential of green steel has been identified as a significant driver of H-DR diffusion, it will not be incorporated into the underlying model for reasons of simplification. As a result, it is assumed that no separate market or additional sales potential exist for green steel, which is therefore treated equally to conventional steel. Within this model, this assumption allows producers to make primary steel production decisions based purely on costs, as the same

[143] Cf. ICAP, 2021, p. 4.

[144] Electricity consumption: 667 kWh/tSteel, grid emission factor (2020): 0.366 $kgCO_2$-eq/kWh, cf. UBA, 2021, p. 9; Vogl et al., 2018, p. 740.

[145] Cf. Agora Energiewende und Wuppertal Institut, 2019, p. 52; Demus et al., 2016, p. 565.

[146] Cf. Baracchini et al., 2019, p. 79; Demus et al., 2016, p. 569; Fidalgo et al., 2015, p. 279.

[147] Shaft furnace electricity consumption: 322 kWh/tSteel, EAF with direct reduced iron: 753 kWh/tSteel (+408 kWh/tSteel compared to secondary steelmaking), cf. Toktarova et al., 2021, p.19; Vogl et al., 2018, p. 740.

[148] Electricity consumption: 50.1 kWh/kgH_2, H_2 consumption: 51 kg/tSteel, emission grid factor: 0.366 $kgCO_2$-eq/kWh, cf. Brändle et al., 2020; UBA, 2021, p. 9; Vogl et al., 2018, p. 739.

[149] Own calculations as described above. The initial value of BF-BOF emission intensity results from harmonizing data on emissions and production volumes, which are obtained from different sources, cf. DEHst, 2014-2021; Worldsteel, 2009-2022.

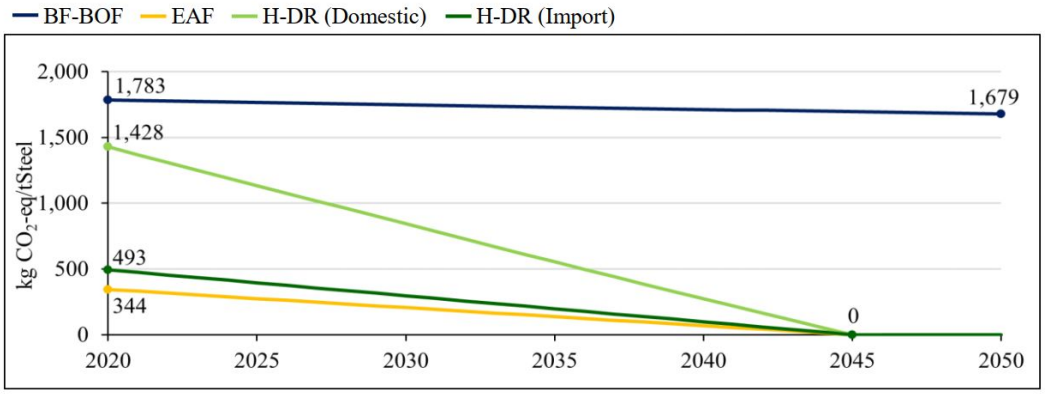


Figure 8: Forecast of emission intensities for the observed production methods.[149]

expected revenue exists for all alternatives. Furthermore, it is assumed that steel producers have perfect information on expected production costs within their planning horizon.

In addition to production costs, the planning horizon of steel producers will serve as the key input for the point in time when the phase-out of conventional steel production is initiated. For steel producers, the decision to start such a transition has far-reaching consequences, as it would entail long-term investments and affect large proportions of total production volume. Consequently, a decision of this type will influence companies' long-term success and can therefore be classified as a strategic decision.[150] The planning horizon for strategic decisions strongly varies between industries and companies. General information is provided by Paul (2015), who considers the strategic planning horizon at more than five years. Weber, Kabst, and Baum (2018) state a similar view, indicating a time horizon of typically five to ten years.[151] To acknowledge the long-term character of steel production, the upper limit of this range will serve as the base assumption of the planning horizon. Thus, the following decision rule is derived: If the expected average costs of H-DR production over the next ten years fall below the analogous costs of BF-BOF production, steel producers initiate hydrogen-based production.

Diffusion Dynamics

In his seminal book Diffusion of Innovations, Everett M. Rogers described innovations as "(...) an idea, practice, or object that is perceived as new by an individual or other unit of adoption"[152], which classifies hydrogen-based steel production as such due to its novelty for steel producers.

An integral part of Rogers' theory is that the cumulative adoption of an innovation follows an s-shaped distribution over time,[153] a proposition that has taken a dominant stance in the diffusion theory of innovation today. Such development is evident, for example, in the diffusion of wind energy, which follows an s-shaped path in many countries.[154] In the German steel industry, such patterns have been observed as well: Arens and Worrell (2014) analyzed historical diffusion dynamics and characterized the diffusion of various technologies such as basic oxygen furnaces or continuous casting machines as s-shaped.[155] Based on these findings, Arens et al. (2017) modeled the future diffusion of H-DR production in Germany with an s-shaped progression in a subsequent study.[156] This assumption is adopted in this thesis.

Different frameworks exist for modeling diffusion processes. A popular approach to model s-shaped diffusion curves can be found in the logistic function, which is described by the following equation:[157]

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}, \tag{4}$$

where L describes the maximum value, x0 the midpoint, and k the function's slope.

To implement the diffusion dynamics according to the logistic function, L equals one in all scenarios, representing complete diffusion at the end of the respective period. However, the decisive factor is the variation of x0 and k, which define the duration until total diffusion and the adoption rate. The determination of these is described below.

Because H-DR technology is still at the beginning of its extensive global application and no historical data are available, determining a realistic diffusion period is subject to significant uncertainties. Nevertheless, to obtain plausible input values, predictable diffusion periods will serve as proxies to derive a realistic time span. A collection of these proxies is provided in the following table:

The most significant problem associated with these observations lies in a lack of representation for the German steel

[150] Cf. Bea & Haas, 2017, p. 327.
[151] Cf. Paul, 2015, p. 165; Weber et al., 2018, p. 92.
[152] Rogers, 1983, p. 11.
[153] Cf. Rogers, 1983, pp. 242-243.
[154] Cf. Davies & Diaz-Rainey, 2011, p. 1235.
[155] Cf. Arens & Worrell, 2014, pp. 972-973.
[156] Cf. Arens et al., 2017, pp. 87-89.
[157] Cf. Sidorov et al., 2021, p. 102.
[158] Cf. ArcelorMittal, 2021, p. 43; Arens et al., 2017, p. 87; H2GS, 2022; Pei et al., 2020, p. 10; Salzgitter AG, 2022b; Thyssenkrupp, 2021, p. 67; Thyssenkrupp, 2022.

Table 4: Proxies for the diffusion period of H-DR production.[158]

Source	Location	Diffusion Period	Description
H_2 Green Steel	Sweden	6 Years (2024-2030)	H-DR introduction until large-scale availability of five million tons per year.
Salzgitter AG	Germany	8 Years (2025-2033)	H-DR introduction until the complete transformation of primary production.
HYBRIT	Sweden	15 Years (2025-2040)	Introduction of H-DR demonstration plants until the comprehensive transformation of industrial plants.
Arens et al. (2017)	Germany	20 Years	Time until total diffusion in Germany, based on historical diffusion periods of other steelmaking technologies.
thyssenkrupp	Germany	20 Years (2025-2045)	Introduction of H-DR until fully carbon-neutral steel production.
ArcelorMittal	Europe	25 Years (2025-2050)	Introduction of H-DR until the Group's steel production is completely carbon-neutral throughout Europe.

industry as a whole. For example, the data from H_2 Green Steel and Salzgitter AG only describe the transformation of a single company, which suggests a more extended period for the entire industry. In the case of thyssenkrupp and ArcelorMittal, only general company targets in the form of carbon-neutral steel production by 2045 and 2050 exist for determining the end of the diffusion period as the companies' H-DR projects lack such precise information. The most realistic specification is presumably found in HYBRIT's project plan considering a diffusion period of 15 years. This outlook describes a transformation across several companies and covers large parts of Swedish steel production. The indication of 20 years by Arens et al. (2017) might serve as a lower bound, as this figure was derived from historical technology diffusions driven purely by efficiencies. Thus, it probably does not fully reflect external factors such as potential political subsidies, emission reduction pressures, or the threat of losing production capacities. Acknowledging the significant uncertainty inherent in this factor, a diffusion period ranging between 10 and 20 years is adopted as being the most predictable at the current time.

For modeling the slope of the diffusion curve, data from Arens et al. (2017) is referred to. They concluded that 5% of H-DR diffusion is reached after three years, i.e., 15% of the total diffusion period.[159] This property is applied proportionally to all underlying diffusion curves. The resulting dynamics are illustrated in the Figure 9.

The input values compiled so far can thus be combined to generate a diffusion scenario. Within this initial projection, the total production volume and the split between primary and secondary production are kept constant at 2021 levels.

4.1.3. Potential Effects on the Production Volume

As constant production volume is assumed in the previous step, this step contains an approach to model effects on the production volume of the German steel industry. The boundary of this model is implemented by emissions budgets for German steel production over the observed period, derived from the German Climate Change Act.

In addition to the increased complexity, more detailed assumptions must be established to provide a framework for consistent modeling. The underlying logic is listed below:

- German steel producers cannot exceed the stated emission budgets;
- If total emissions exceed the emission budgets, producers must reduce their production volume accordingly;
- When forced to reduce their production volume due to emission restrictions, producers maximize their output by expanding less emission-intense secondary production, reducing BF-BOF production only;
- Once BF-BOF production volume is reduced due to emission restrictions, it is lost and cannot be recovered;
- As no separate market for green steel exists, production of it cannot develop additional markets, hence only the remaining BF-BOF capacity can be transformed into H-DR production;
- The reduction of production volumes does not affect production costs and H-DR diffusion dynamics.

Before introducing the input variables underlying this approach, it is to be noted that the presented logic will most likely not be implemented in this form in reality. Instead, it is opposed to the current policy framework, as the foreseen distribution of free allowances until at least 2030 was designed to preserve German steel production. Strict enforcement of emission budgets, at least while free allowances are still being issued, therefore does not seem logical. In the context of these circumstances, this analysis aims to examine the degree of incompatibility of German steel production with the

[159] Cf. Arens et al., 2017, p. 87.
[160] Own calculations.

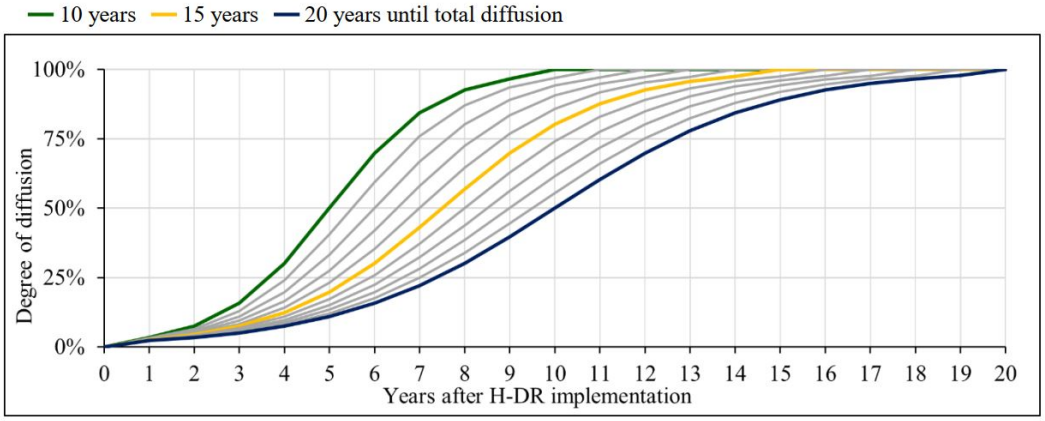


Figure 9: Underlying diffusion dynamics.[160]

applicable emission targets. The vulnerability of the steel industry to migration effects, which would result from a strict implementation of emission restrictions without further support measures, is also to be investigated.

Considering the emission budgets set by politics in the Federal Climate Change Act, it becomes apparent that these were only defined with limited granularity. Targets specific to the steel industry are not indicated, yet they have been specified for the industrial sector as a whole. As explained in Chapter 3.1, German steel production was responsible for a substantial share amounting to 28% of total industrial emissions in 2020, making a restriction by these superordinate targets reasonable. Consequently, the industrial emission budgets are applied proportionally to the emissions of the steel industry. Until 2030, specific annual limits exist. Afterward, emissions are interpolated linearly to the target of carbon neutrality in 2045. This results in the following emission budget for German steel production in the coming decades:

The production volumes are linked to the emission intensity forecasts. The product of these two factors constitutes the emissions generated and must therefore not exceed the restrictions evident in Figure 10.

The final input variable is the development of secondary steelmaking, which is expected to expand when steel producers are forced to reduce their production due to emission restrictions. This logic is based on the assumption that secondary production is likely to become more dominant in the future, as discussed in this thesis's qualitative part. For the expansion of secondary production, a framework is provided in a joint study conducted by the German Steel Institute and the Boston Consulting Group. These concluded that future European secondary production volumes will primarily be limited by the availability of the required steel scrap. Until 2050, the study predicts an annual increase in this availability of 0.9% - an assumption adopted for this model.[162]

Now that all variables have been introduced, Table 5 provides an overview of varying factors, including the values each can assume in the underlying model. Combining these variables forms the basis for extracting the final scenarios, which will be carried out in the next section.

4.2. Scenario Extraction

In principle, it would be possible to create 108 different combinations from the drivers presented in Table 5. However, it is necessary to extract individual combinations to illustrate pivotal interrelationships in the observed environment as desired. Insights from scenario planning literature are relied upon for this purpose.

Amer, Daim, and Jetter (2013) provide a comprehensive review of scenario planning literature in which they compare and discuss results from numerous studies. Thereby, the authors identified particularly two properties, which are repeatedly regarded as fundamental for the credibility of scenarios: plausibility and consistency. The plausibility criterion refers to the fact that an occurrence of scenarios should be realistic. The consistency criterion addresses the need for the combinations of individual driver values to follow a clear logic and not be contradictory.[163] Considering the number of generated scenarios, most researchers favor the development of three to five scenarios to achieve the best trade-off between manageability and insight quality.[164]

For the final scenario selection, the role of the policy framework serves as the foundation. To ensure plausibility and internal consistency, the aim is to derive coherent combinations of the other drivers based on the policy framework. Since the continuation of the current free allowances regime has been classified as the likeliest, two scenarios are created for this variant. Furthermore, one scenario each is subject to an early phase-out and the complete absence of free allowances. The four scenarios that have proven to be particularly meaningful are presented below.

[161] Actual emission data included until 2020, cf. DEHst, 2014-2021; UBA, 2022. Values after 2020 are based on specified targets in Federal Climate Change Act, Section 3 & Annex 2.
[162] Cf. Woertler et al., 2013, pp. 33-35.
[163] Cf. Amer et al., 2013, p. 37.
[164] Cf. Amer et al., 2013, pp. 32-33.

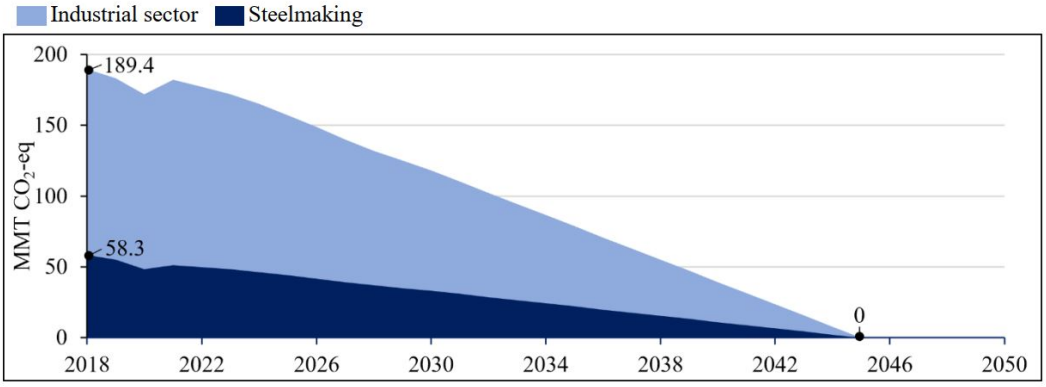


Figure 10: Emission budgets derived from the German Climate Change Act.[161]

Table 5: Summary of drivers included in the model.

Variable	Potential Values	Notes
Policy framework	▲ No free allowances ► Early phase-out ▼ Current policy	Lead indicator for each scenario.
CO_2 prices	▲ NZE ► APS ▼ STEPS	Adopted from IEA(2021).
Hydrogen source	▲ Import ▼ Domestic production	Domestic production mainly suitable as transition technology.
Hydrogen costs	▲ Optimistic ▼ Baseline	Adopted from Brändle et al.(2020).
Diffusion speed	▲ 10 years ► 15 years ▼ 20 years	Extracts from the derived period of 10 to 20 years.

▲►▼ High/ medium/ low support of H-DR diffusion

4.2.1. Scenario 1: Current Policy – Downside (CP-D)

This scenario is subject to the current free allowances regime, which means that steel producers receive free allowances for all emissions until 2030, followed by a phase-out until 2040. Additionally, this scenario has the objective of reflecting foreseeable downside risks for H-DR diffusion. It is assumed that policy support measures exceeding the current measures will not be realized, resulting in a sluggish increase in CO_2 prices according to the Stated Policies Scenario by the IEA. Furthermore, no subsidies are provided for green hydrogen projects, negatively affecting their costs. Due to missing cost pressures and a lack of policy support, steel producers show little willingness to transform, resulting in slow uptake of H-DR technology.

4.2.2. Scenario 2: Current Policy – Baseline (CP-B)

This scenario likewise relies on the currently anticipated free allowance policy for steel production but draws a more positive picture of future developments. A higher willingness of politicians to support sustainable projects leads to a faster increase in CO_2 prices following the Announced Pledges forecasts of the IEA. However, measures to support green hydrogen production remain absent, leaving them unchanged compared to the first scenario. Nevertheless, steel producers face increased transformation pressure mainly due to rising emission compensation costs and implement H-DR production faster than would be the case through pure efficiency gains.

4.2.3. Scenario 3: Early Phase-Out (EPO)

Scenario 3 assumes that the current free allowances regime is modified into a phase-out between 2026 and 2035. This industry-specific measure arises from a strong willingness of policymakers to accelerate the energy transition, leading to a rapid increase in CO_2 prices as anticipated in the Net Zero Emissions By 2050 Scenario by the IEA. Green hydrogen projects are being promoted, resulting in an improved cost outlook. As this exerts intensified cost pressure on con-

ventional steel production and simultaneously increases the attractiveness of H-DR production, steel producers show a high willingness to transform, which manifests in a short diffusion period.

4.2.4. Scenario 4: Best Case (BC)

This scenario is based on the EPO Scenario but additionally displays the consequences of an absence of free allowances over the entire observation period. Even though this assumption is somewhat unrealistic, this scenario represents the best possible case and will thus mainly serve as a benchmark.

The following table provides an overview of the specific input values for each scenario:

4.3. Results

In this chapter, the results obtained through the developed model are presented. After an excursus on the potential of domestically produced hydrogen, the analysis of the selected scenarios follows. This analysis is structured in line with the previous chapter: After considering the results obtained from the tipping point analysis, an examination of the corresponding diffusion scenarios follows. Then, effects on production volumes resulting from strict adherence to emission budgets are investigated. Nevertheless, the general informative value of the developed model must first be assessed.

The very nature of scenarios implies essential elements that must always be taken into account when analyzing results of this kind: Scenarios are not exact predictions of what the world will look like tomorrow. Instead, their purpose is to create a broad perspective on fundamental trends and uncertainties.[165] Complex aspects are ordered and woven into "(...) coherent, systematic, comprehensive, and plausible"[166] stories to map the range of potential alternatives.[167] Such a simplified structuring of complex relationships is also subject to this model and is essential to identify the key interdependencies in the underlying long-term approach. Moreover, the included input variables are exposed to high uncertainties or do not yet offer any actual data, which also must be considered. Fischedick et al. (2014) noted the consequence of such limitations during a similar approach. The researchers argued that absolute numerical values should be assumed to be less meaningful. Instead, the observable relative correlations and the comparison of different scenarios allow the most reliable conclusions to be drawn.[168] This reasoning is adopted for the results presented here.

4.3.1. Potential of Domestic Hydrogen Production

As illustrated in Table 6, all extracted scenarios assume that the hydrogen required for steel production will be imported and that domestically produced hydrogen will thus be of minor importance. Therefore, before presenting the scenario analysis results, the rationale for this selection is stated.

As already explained in Chapter 3.2.4, the domestic production of hydrogen is associated with considerable restrictions compared to hydrogen imports, which led to the conclusion that this variant could primarily draw potential as a transition technology until sufficient hydrogen import infrastructure is available. However, within the developed model, correlations became visible that seem to limit this potential and led to the exclusion of domestic production from all scenarios.

Comparing the impact of the different hydrogen sources on the cost projection of H-DR production, domestic production is associated with significantly higher costs than imports. This disadvantage is illustrated in Figure 11 for the BC Scenario: There, the additional costs of domestic hydrogen production lead to a delay of five years until cost equality to the BF-BOF is achieved. Within the underlying mechanisms, this has the consequence that domestically produced hydrogen does not offer sufficient incentives in any of the scenarios to initiate the H-DR roll-out in the critical phase until around 2030,[169] in which it could serve as a substitute for imported hydrogen. Even within the BC Scenario, where the assumption of strong policy support favors earlier hydrogen imports, implementing H-DR steelmaking using hydrogen produced in Germany is only feasible from 2030 onwards.

One significant factor influencing this development is the contradictory development of domestic hydrogen production costs compared to CO_2 prices. As illustrated in Figure 7, higher CO_2 prices lead to higher hydrogen production costs, driven by the compensation costs for the generated emissions. Thus, while high CO_2 prices in principle foster the early implementation of H-DR technology by raising the cost of conventional steel production, they simultaneously disadvantage domestic hydrogen production. Within the defined framework, it can therefore be concluded that domestic hydrogen production in Germany does not offer enough incentives to unfold its potential as a transition technology without further support.

4.3.2. Tipping Point Analysis

This section presents the results obtained from the analysis of the four extracted scenarios. First, the tipping point analysis is discussed, which considers the cost forecasts of the observed steelmaking technologies until 2050.

Figure 12 visualizes the breakdown of production costs by cost factor for each production method in 2022, representing the starting point of the projections.

This analysis shows that raw material and energy costs account for the largest part of production costs in all methods. In 2022, these two cost factors account for 56% of total costs

[165] Cf. Schoemaker, 1995, p. 28.
[166] Coates, 2000, p. 116.
[167] Cf. Hiltunen, 2009, p. 151.
[168] Cf. Fischedick et al., 2014, p. 567.
[169] This period is derived from the plans of the European Hydrogen Backbone which include establishing initial European hydrogen infrastructure starting in 2030, cf. Wang et al., 2020, p. 4.
[170] BC scenario underlying.

Table 6: Input values underlying the selected scenarios.

Driver \ Scenario	Current Policy – Downside	Current Policy – Baseline	Early Phase-Out	Best Case
Policy framework	▼ Current policy	▼ Current policy	► Early phase-out	▲ No free allowances
CO_2 prices	▼ STEPS	► APS	▲ NZE	▲ NZE
Hydrogen source	▲ Import	▲ Import	▲ Import	▲ Import
Hydrogen costs	▼ Baseline	▼ Baseline	▲ Optimistic	▲ Optimistic
Diffusion speed	▼ 20 years	► 15 years	▲ 10 years	▲ 10 years

▲ ► ▼ High/ medium/ low support of H-DR diffusion

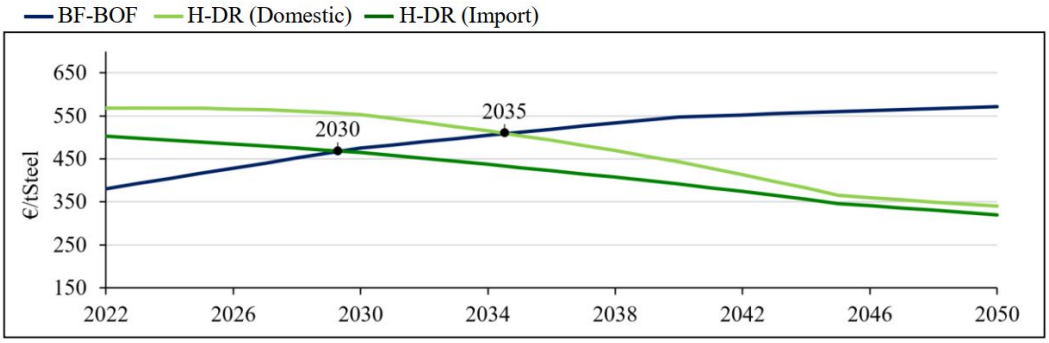


Figure 11: Impact of domestic hydrogen production on H-DR cost forecast.

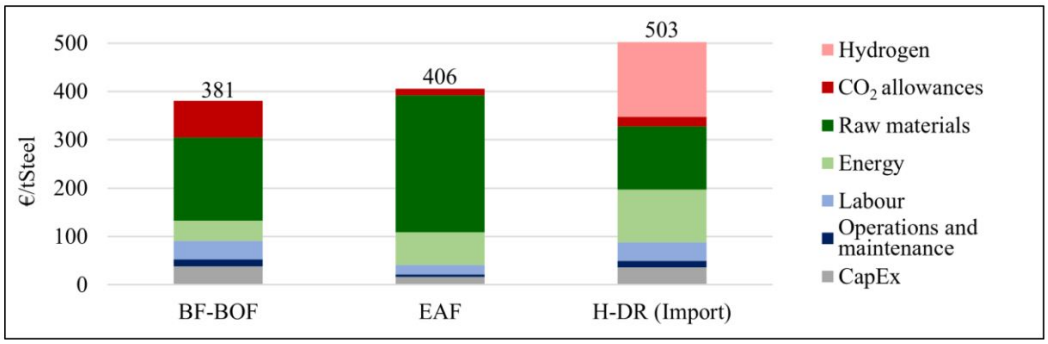


Figure 12: Production cost breakdown of the observed methods (2022).[170]

within BF-BOF production and about 87% in secondary production. The increased exposure of secondary production is caused by the high consumption of the relatively expensive steel scrap and the high electricity demand of this method. For H-DR production, this share, including hydrogen costs, is also responsible for the largest part of total costs, amounting to 78% in 2022. Since this illustration is based on the BC Scenario, the production costs already include compensation expenses for caused emissions, a factor that is not yet present in reality. However, even in this analysis, the BF-BOF method turns out to be the substantially more economical primary production method as the H-DR method causes about 32% more costs. If the costs of CO_2 allowances are disregarded, this share even rises to almost 59%, demonstrating the enormous cost differences between the two primary production methods at the beginning of the observation period.

Based on the described production costs, various combinations of CO_2 prices and hydrogen cost developments are underlying, affecting the initial cost gap differently. The resulting tipping point analysis is provided in Figure 13.

It becomes evident that mainly the costs of BF-BOF and H-DR production change significantly in all scenarios. Due to its low emission intensity and as no hydrogen is required, the changes in EAF production costs are much less pronounced.

Within the CP-D Scenario, relatively limited effects on individual production costs are evident and result in H-DR pro-

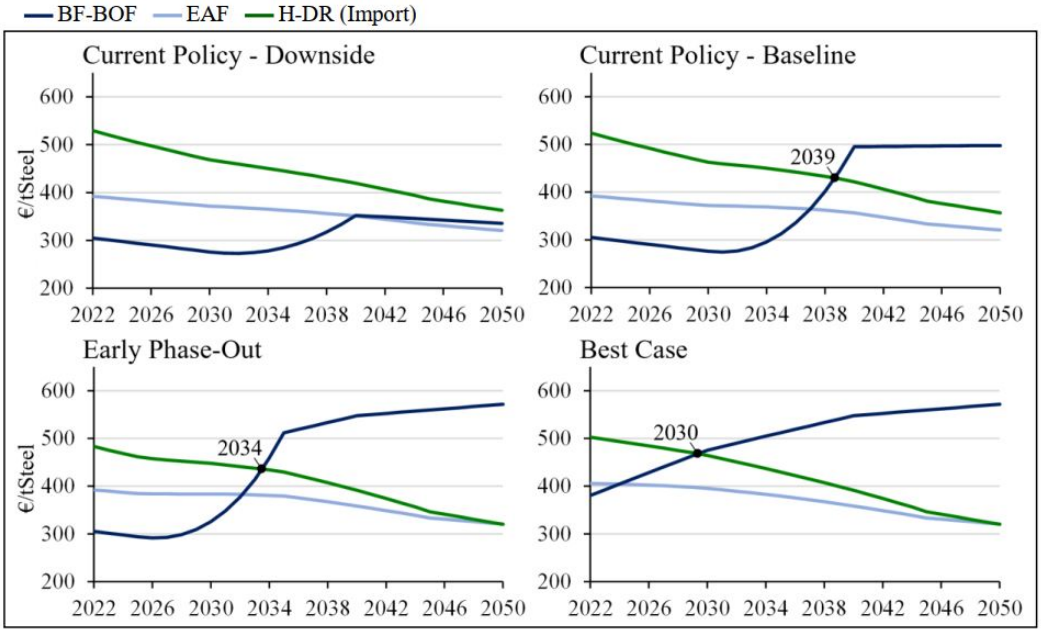


Figure 13: Tipping point analysis of the selected scenarios.

duction imposing higher costs than BF-BOF production over the entire period. Low CO_2 prices emerge as a major cause, as even the complete absence of free allowances from 2040 onwards does not result in a sufficient increase in BF-BOF costs to achieve cost advantages of H-DR production. This differs in the CP-B Scenario: A significant increase in BF BOF production costs is apparent, driven by higher CO_2 prices combined with the phase-out of free allowances. This development ultimately leads to reaching the tipping point in 2039.

Within the EPO and BC Scenario, a stronger trend in production costs emerges. Due to lower hydrogen costs and rapidly increasing CO_2 prices, the costs of the individual methods converge faster and result in cost advantages of H-DR production starting in 2034 and 2030, respectively. The significant effects caused by a phase-out of the free allowances in combination with high CO_2 prices can be observed particularly. While BF-BOF production still has the lowest costs at the beginning of the observation period, these rise sharply in the phase-out period due to their high emission intensity and thus considerably influence the timing of the tipping point. Because of their low emission intensity, EAF and H-DR production costs show only minor changes caused by the phase-out, not affecting their overall trend.

Further insights can be obtained by examining the leverage of the individual cost factors on the tipping point timing. For this purpose, each of them is modified in all scenarios under otherwise constant conditions (ceteris paribus). The alternative values of the cost factors are found within the dimensions defined for the model, summarized in Table 5. For each of these modifications, the triggered shift of the tipping point in years is recorded in Figure 14. Values indicating a change to a point whose location lies outside the period depicted in the model refer to the year 2051. Thus, the given information reads as follows: If the CP-D Scenario would be subject to the development of CO_2 prices according to the NZE instead of STEPS forecasts, the tipping point between H-DR and BF-BOF production would be reached 12 years earlier. Since the CP-D Scenario is not subject to a tipping point within the observation period until 2050, the described change would result in reaching the tipping point in 2039, i.e., 12 years before 2051.

Considering these findings, it becomes apparent that the individual cost factors exert varying degrees of influence on the timing of the tipping point. By far, the greatest leverage is found in the CO_2 prices: Especially the step between the STEPS and APS projections turns out to be impactful and results in a shift of the tipping point by at least twelve years in all scenarios. The step between the APS and the NZE forecasts is of significantly smaller importance and merely causes a shift of one year in the BC Scenario. The factor with the second-largest impact is the underlying policy framework, whose variation mostly causes a shift of four to eight years. Only in the CP-D Scenario its modification does not influence the timing of the tipping point. This lack of stimulus is caused by the factors' underlying dimensions, which only constitute different values up to 2040 since no free allowances are underlying from this point on at the latest. As the tipping point of the CP-D Scenario is far beyond 2040, changes in this factor no longer have any influence. The situation is different for the hydrogen costs: In the CP-D Scenario, a reduction would result in a shift of the tipping point to 2046. Among the other scenarios, this factor modification only results in minor changes.

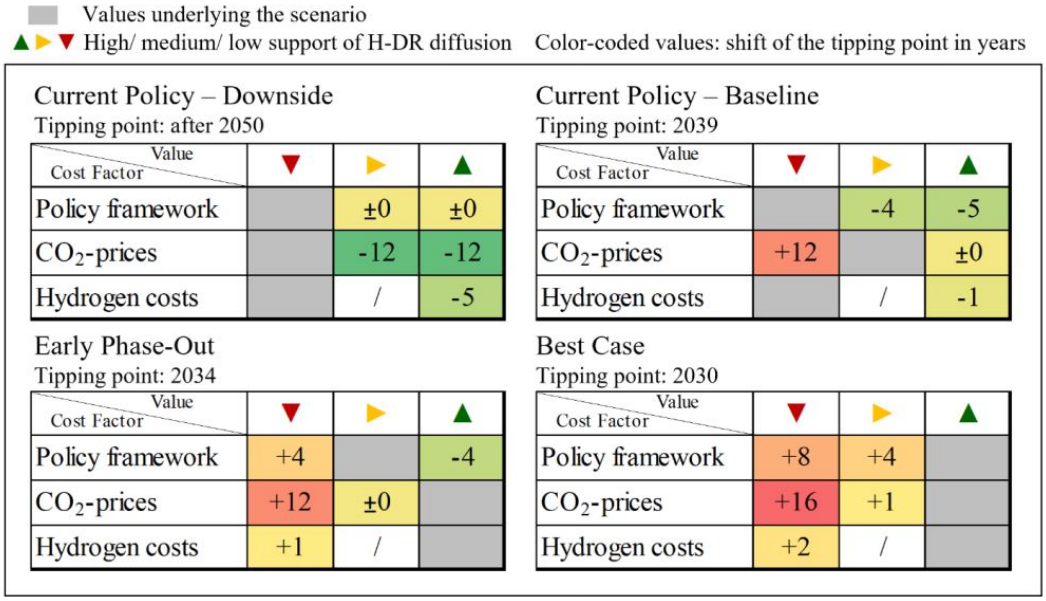


Figure 14: Effects of the individual cost factors on the timing of the tipping point.

4.3.3. Diffusion Scenario of Hydrogen-Based Steelmaking

The analysis of the diffusion scenario is carried out according to key indicators, which primarily include the respective emission profiles and cost projections resulting from the different diffusion dynamics. The production volume is assumed to remain constant at the 2021 level and thus includes the production of 27.98 MMT of primary steel and 12.09 MMT of secondary steel.[171]

Figure 15 provides an overview of the diffusion scenarios of all selected scenarios.

What strikes first are the differences in the starting points of the respective H-DR rollout, which are derived from the tipping point analysis according to the underlying decision rule of producers explained in Chapter 4.1.2. While the BC Scenario would result in the initiation of H-DR production already in 2026, a substantial delay occurs in the other scenarios. However, the CP D Scenario is particularly prominent. There, the H-DR costs exceed the BF-BOF costs too much over the entire period to enable sufficient stimulus for triggering H-DR production. Given the long diffusion period of 20 years underlying the CP-D Scenario, this indicates that a comprehensive transformation will not occur until well into the second half of the century. However, a distinct shift can be identified for the CP-B Scenario. This scenario anticipates an implementation nine years after the BC Scenario. It would thus result in a complete transformation of steel production at the end of the observed period in 2050. Nevertheless, it ranks well behind the BC and EPO Scenarios, including a change in the free allowances regime and a 14 and 10 years earlier completion of the transformation.

[171] Cf. WV Stahl, 2022, p. 1.

Emissions

The emissions profile of each scenario is of fundamental importance to assess their compatibility with the emissions budgets targeted by the federal government. An overview of annual and total emissions resulting from each scenario is provided in Figure 16.

A finding that emerges from the analysis of annual emissions is that no scenario can realize the annual reduction targets of the Federal Climate Change Act in the near future. Even the EPO and BC Scenario do not foresee realizing them until 2035 and 2030. However, their trajectories reveal that H-DR production contributes to vast emission reductions after its implementation: A sharp decline in annual emissions occurs and enables significantly lower emissions than budgeted. The long-term goal of carbon neutrality by 2045 is achieved within both scenarios despite temporary budget overruns.

The analysis of the CP-D and CP-B Scenario reveals far more pessimistic emission trends. Due to the lack of H-DR introduction, the emission savings within the CP-D Scenario are exclusively based on improvements to the established production methods. This results in only a minor reduction of annual emissions by about 12%, from 53.5 to 47.0 MMT CO_2-eq between 2022 and 2050. Consequently, the emissions budgets are vastly exceeded over the entire observation period. The CP-B Scenario is also subject to exceeding the emission targets until 2049. However, the introduction of H-DR technology still facilitates a completely carbon-neutral production from 2050 onwards, five years later than targeted by policy.

Regarding the total emissions, it can be concluded that only the BC Scenario, with total emissions of about 528 MMT CO_2-eq between 2022 and 2050, can remain below the budget, which amounts to 607 MMT CO_2-eq. Although the

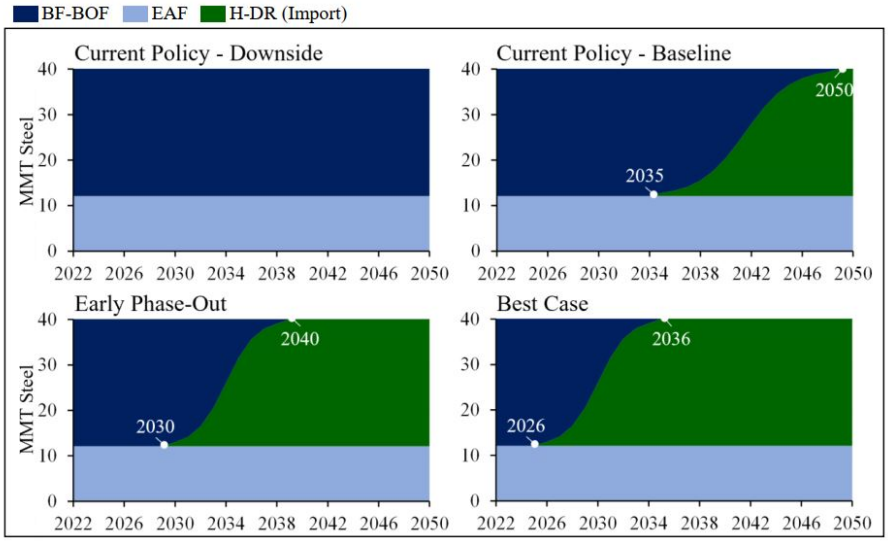


Figure 15: Diffusion dynamics of the selected scenarios.

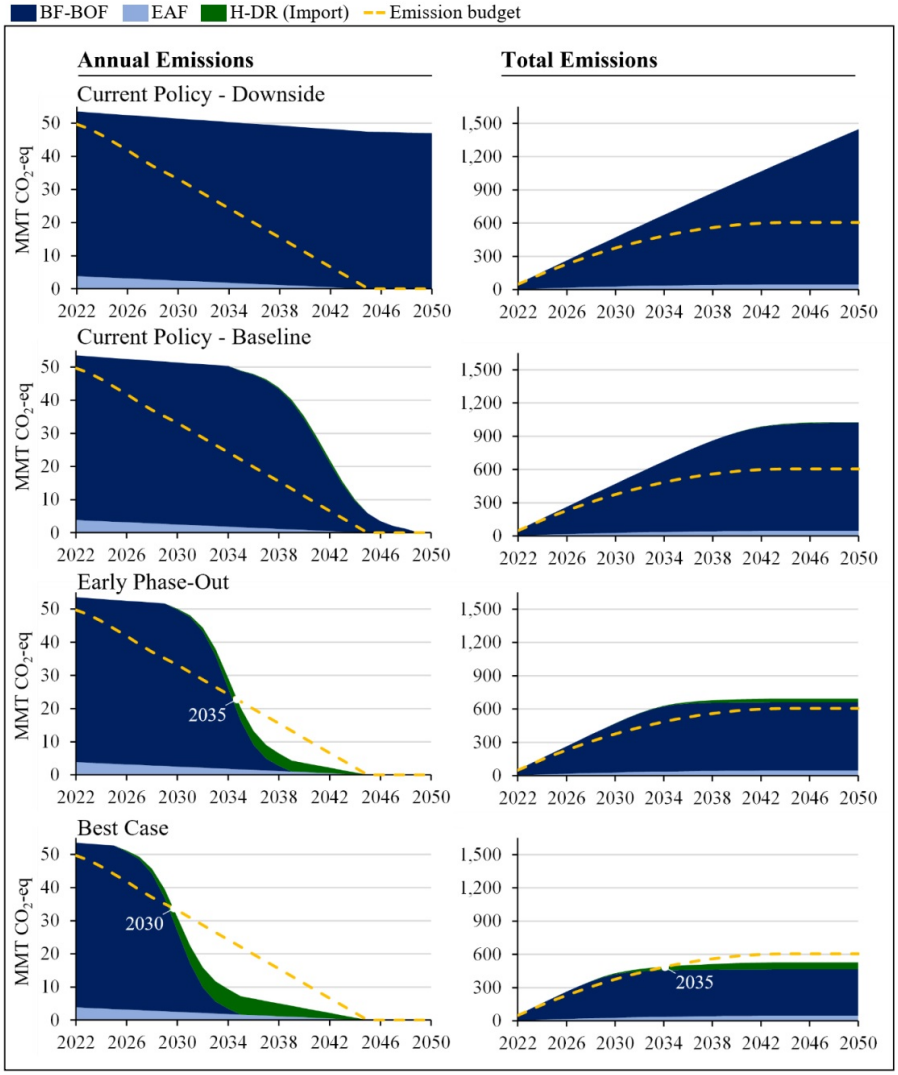


Figure 16: Annual and total emissions of the selected scenarios.

EPO Scenario undercuts annual target emissions beginning in 2035, the subsequent savings are insufficient to offset the previously generated excess emissions. With total emissions of 694 MMT CO_2-eq, it misses the budget by about 14%. Both scenarios, which include the continuation of the current free allowances regime, result in massive overruns of the budget regarding the total emissions caused. With emissions of 1,026 MMT CO_2-eq, the CP-B Scenario already exceeds the budget by 69%. Nevertheless, it raises expectations of a positive development after 2050 since the completed transformation in 2050 ensures that subsequent steel production will be emission-free. The situation is different for the CP-D Scenario: There, the total emissions of 1,447 MMT CO_2-eq exceed the budget by 138% and do not indicate any improvement, as the H-DR implementation has not yet occurred.

Production Costs

The analysis of production costs is based on calculating the average production costs in each scenario. The average costs are composed of the different production methods applied, weighted by their shares in total production volume. Additionally, the development of average costs is displayed for the case where no H-DR adoption takes place, and no allowances have to be purchased to compensate for emissions. Figure 17 provides an overview of the underlying development.

Examining the average costs shows that these are subject to a temporary increase in each scenario, followed by a reduction. It can be observed that the earlier the scenario conditions stimulate the H-DR rollout, the higher the average production costs. Averaging at 391 € /tSteel over the entire period, the BC Scenario features the highest costs reaching a maximum value of 447 € /tSteel in 2030. The EPO Scenario follows in second place and includes average costs of 358 € /tSteel. Its highest value of 432 € /tSteel is reached in 2035, the year the phase-out of free allowances is terminated. The CP-B and CP-D Scenario appear even more favorable, with average costs of 351 and 325 € /tSteel, respectively. The maximum values of 438 and 352 € /tSteel for both are found in 2040, which again is the first year purchased allowances must fully compensate for the generated emissions. The influential role of free allowances becomes evident from an overall perspective: In all scenarios involving a phase-out, an upward trend of the average costs along it is evident. Furthermore, the BC Scenario, which is not subject to any free allowances, entails significantly higher costs than the other scenarios from the very beginning.

Although CP-D Scenario turns out to be the scenario with the lowest average costs, the developments at the end of the observation period and their outlook prove to be particularly informative. After reaching their maximum average costs, all scenarios, which include transforming to H-DR production, are subject to a stronger cost reduction trend than the CP-D Scenario. The forecasts even suggest that the average costs of the BC and EPO Scenario undercut those of the CP-D Scenario in 2046, making them the least expensive variants from that point forward.

However, comparing the cost curves of the scenarios to the costs of BF-BOF production that is not subject to any emissions allowance costs, it becomes clear that the CO_2 costs burden steel production with long-lasting disadvantages in any case.[172] At an average of 285 € /tSteel, this option offers by far the lowest costs over the entire period. Furthermore, even towards the end of the observation period, only minimal convergence of the scenarios involving an H-DR transformation is evident. Conventional steel production, which is not subject to emission compensation payments, can therefore be characterized as the least expensive production alternative.

Costs of Potential Subsidies

Another interesting perspective derived from the analysis of production costs is found in the payments required to support H-DR production in each scenario. This analysis roughly follows the mechanism of a CCfD as explained in Chapter 3.4.2. The subsidy amount is defined as the difference in production costs between H-DR and BF-BOF production per ton of steel, which is then offset against the targeted H-DR production volume for each year. Thus, if H-DR costs exceed BF-BOF costs, expenses will be incurred if H-DR production is realized. Since perfect information about the course of production costs is available within this model, the subsidy amount is calculated individually for each year instead of defining a strike price over a prolonged period, as would be the case in a real setting. The BC Scenario will serve as a benchmark in the underlying case: The costs incurred to align the H-DR diffusion with the BC are calculated for each scenario. It is important to note that this calculation refers to the BF-BOF costs of the respective scenario, including costs for CO_2 allowances. Conclusions on the required subsidy volume to achieve cost parity with foreign producers, which are not subject to CO_2 prices, are not directly feasible. The respective payments are visualized in Figure 18.

In each scenario, it becomes evident that additional payments amounting to several billion euros would be required to achieve the targeted production volume. Significant differences between the individual scenarios are apparent. The EPO Scenario involves total payments of € 8.1 billion, around 20% of the German steel industry's revenue in 2019[173], spread over eight years after the H-DR introduction in 2026. The highest annual payments are required in 2031 at € 1.88 billion, and after 2034, no additional costs are generated as the H-DR production costs fall below those of BF-BOF production. A large step is evident in the CP-B Scenario, which, at costs of € 31.7 billion, results in nearly a quadruple of required funding compared to the EPO Scenario, mainly driven by higher annual payments, as shown in Figure 18. At the beginning of the H-DR introduction, the annual costs increase rapidly and reach their maximum of € 4.4 billion in 2033. A similar development is underlying the CP-D Scenario: The required annual payments also reach their maximum in 2033 at € 4.53 billion. Since, in this sce-

[172] Represented by the dashed line in Figure 17.

[173] Revenue of the German steel industry in 2019: € 39.8 billion, cf. WV Stahl, 2021b, p. 13.

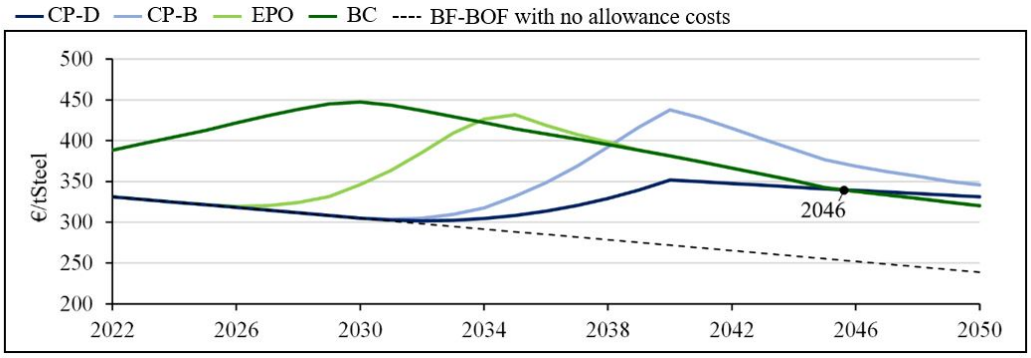


Figure 17: Average production costs of the selected scenarios.

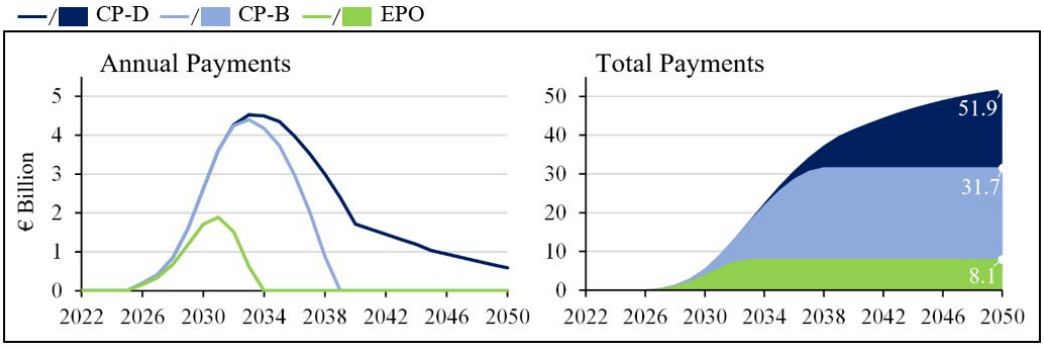


Figure 18: Required payments to achieve the diffusion of the Best Case Scenario.

nario, the H-DR costs exceed those of BF-BOF over the entire period, payments are necessary throughout all years to promote H-DR production. These amount to € 51.9 billion in total - 130% of the industry's turnover in 2020.

A relationship between the annual payments and the phase-out of the free allowances becomes apparent and is particularly evident within the CP-D Scenario. There, the end of the phase-out in 2040 abruptly causes the required annual payments to decrease at a slower rate. This deceleration illustrates that free allowances result in higher annual payments and that their phase-out contributes to rapidly reducing the required payments. This observation is confirmed when considering the primary impact of free allowances: By lowering the cost of BF-BOF compared to H-DR production, higher payments are necessary to achieve cost parity.

4.3.4. Potential Effects on the Production Volume

When evaluating the results in the previous section, it must always be taken into account that these are based on the assumption of constant production volumes and a constant split between primary and secondary production. Effects of excessive emissions or additional costs on the production volume are not reflected and therefore represent a simplification of reality. This analysis step follows an approach to evaluating potential effects on production volumes caused by strict enforcement of the imposed emission budgets. Again, the production volume of the German steel industry in 2021 serves as the starting point.

Figure 19 illustrates the evolution of production volumes for each scenario resulting from consistent adherence to the annual policy emission budgets.

Since the Federal Climate Change Act considers emission reductions not feasible with conventional production methods already for 2022, each scenario entails a reduction in primary production capacities from the beginning. Thus, steel producers also begin expanding secondary production from the start in all scenarios, limited by steel scrap availability as defined in the underlying logic.

The analysis of the CB-D Scenario again indicates that BF-BOF production is incompatible with the targeted emission reductions. There, the lack of H-DR implementation results in losing the total primary production capacity to enable achieving the emission targets. Thus, the German steel industry consists exclusively of secondary production at the end of the period and has shrunk massively overall. The loss of 27.98 MMT of primary steel is offset by an additional production of 3.59 MMT of secondary steel, reducing German steel production by 24.39 MMT to 39% of its initial size.

Within the CP-B Scenario, the introduction of H-DR allows at least parts of the primary production capacity to be transformed before becoming irreversibly depleted due to emission restrictions. Nevertheless, because of the relatively late H-DR roll-out, this scenario also entails losing most BF-BOF production. About 60% of primary steel production, corresponding to 16.73 MMT Steel, are dismantled until 2037. Considering the additional secondary steel pro-

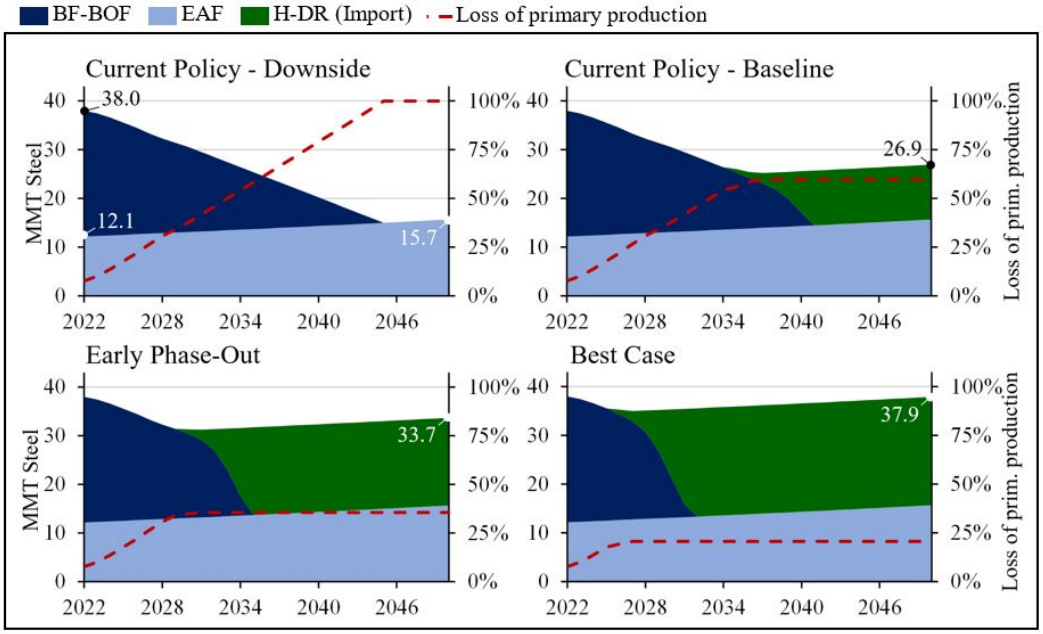


Figure 19: Effects of strict compliance with emission budgets on production volumes.

duction, which is identical in each scenario, a loss of 33% of total steel production is implied.

In the two more optimistic scenarios, the majority of primary steel production can be maintained and transformed. In the EPO Scenario, the reduction in primary production until 2031 results in losing 9.98 MMT, around 36% of its initial volume. This loss equates to a 16% reduction in total production after taking the expansion of secondary production into account. Within the BC Scenario, the required reduction until 2027 is associated with primary steel losses of 5.78 MMT, corresponding to 21% of the initial production. However, this loss can be compensated by additional secondary capacities to a large extent. As a result, the loss of total production volume adds up to 5% of its initial volume.

In summary, each scenario would be associated with a decline in overall production if annual emissions budgets were strictly enforced. Nevertheless, the severity of the triggered reduction differs greatly between the scenarios. Furthermore, the near-term emissions budgets appear to pose the greatest challenges, as these cannot be achieved in any of the scenarios, thus implying the largest output reductions.

4.3.5. Sensitivity Analysis

Since the underlying model is designed to forecast future developments, some of the contained variables are subject to considerable uncertainty. To identify the contribution of individual input values to the prevailing uncertainty and thus to better understand their role within the model, a suitable tool is found in the sensitivity analysis.[174] For this purpose, individual values are varied while holding all other constant and analyzing the impact on the model output. The sensitivity analysis will be performed from two perspectives as the model produces various outputs. First, the sensitivity towards the most significant cost factors will be tested by examining their impact on the timing of the tipping point, followed by an analysis of parameters underlying the diffusion dynamics. The foundation for the sensitivity analysis consists of the CP-B Scenario.

For H-DR production, the most significant cost factors are the costs of electricity, iron ore, and hydrogen. In 2039, the year of the tipping point in the underlying scenario, these account for about 72% of the total production costs. Accordingly, the costs of scrap, CO_2 allowances, and iron ore are the most relevant factors for BF-BOF production, accounting for 67% of total costs in 2039. Table 7 shows the results of the sensitivity analysis. The variation is applied to all cost factor values up to 2050, the color-coded figures indicate the shift of the tipping point caused by the variation.

In principle, it can be observed that the location of the tipping point is relatively robust to uncertainties in the cost factors. For example, massive changes in iron ore costs exert no influence at all, as these are subject to both production methods and cause a parallel shift of the cost curves. Electricity and steel scrap prices likewise demonstrate limited influence, although these each affect only one of the two production methods. However, the model shows increased sensitivity to hydrogen costs and especially CO_2 prices since significant shifts in the tipping point occur when these deviate. Although these two factors are already included as critical indicators in the scenario analysis and different values were

[174] Cf. Saltelli, Tarantola, Campolongo, & Ratto, 2004, p. 45.

[175] Values indicating a shift to after the observation period refer to the year 2051. CP-B Scenario underlying.

Table 7: Sensitivity analysis of major cost factors.[175]

Variable \ Variation	-80%	-60%	-40%	-20%	+20%	+40%	+60%	+80%
Iron ore	±0	±0	±0	±0	±0	±0	±0	±0
Electricity	-2	-1	-1	±0	+1	+1	+1	+2
Hydrogen	-3	-2	-1	±0	+1	+1	+2	+5
Scrap	+1	+1	+1	±0	±0	±0	±0	-1
CO_2 prices	+12	+12	+6	+1	±0	-1	-1	-2

Color-coded values: shift of the tipping point (2039) caused by variation in years.

considered for them, the major impact of the uncertainty underlying these factors on the model output must always be taken into account.

Considering the assumptions underlying the modeling of the H-DR diffusion, decisive input values can be found in the H-DR initiation year and diffusion speed. Since H-DR initiation is derived from the cost developments of H-DR and BF-BOF production and the planning horizon of producers, conclusions on these can simultaneously be drawn. The analysis of the diffusion rate is justified by its subjection to major uncertainties resulting from the lack of historical values or distinct proxies. The total emissions generated are applied as model output, whose change caused by the variation of the variables is observed. Table 8 presents the results of the sensitivity analysis. The variation describes the shift of the H-DR initiation or the shortening or lengthening of the time until complete diffusion is achieved in years, the colored values describe the triggered change in cumulative emissions until 2050.

A significant influence on the model output can be detected for both observed variables. However, in direct comparison, the deviations in the H-DR initiation are associated with approximately double the impact on the selected model output. If these findings are combined with those of the first step of the sensitivity analysis, this again illustrates the enormous influence that is in particular exerted by uncertainties in CO_2 prices throughout the model, which they exercise by influencing the timing of the tipping point.

4.4. Discussion

After the scenario analysis results were presented in the previous section, these will now be discussed in the research questions' context. Furthermore, potential limitations underlying the results will be considered. The discussion is structured as follows: First, observable correlations are highlighted at a general level, then the current political framework is assessed, and finally, recommendations for policymaking are derived.

4.4.1. Limitations of the Developed Model

The limitations of the proposed model are primarily found in its simplifying assumptions, which may affect the validity of the results. For instance, it was assumed for BF-BOF production to be continued only in its current form. However, additional emission savings could arise from applying BF-BOF production through modifications as a transitional solution. Examples of this can be found in applying CCS technologies or recycling the exiting gas from the blast furnace, as already considered by other papers.[177] The utilization of hydrogen produced with natural gas in the H-DR could likewise offer potential as bridging technology before green hydrogen will be extensively available.[178]

Another limitation is found in the fact that no added value is attributed to green steel compared to conventional steel. As stated in the first part of this thesis, the development of such market potential could significantly influence the future of H-DR production. If, for example, consumers express a greater willingness to pay for green steel, additional incentives for transformation could arise for steel producers beyond purely cost-based decisions as considered in the underlying approach. Thus, this limitation is identified as upside potential that might cause significant shifts in the developed scenarios.

4.4.2. General Findings

The modeling aimed to depict realistic scenarios for the diffusion of hydrogen-based steelmaking subject to various policy frameworks surrounding the German steel industry. It becomes clear that the future development of steelmaking is only vaguely foreseeable, reflected in significant differences between the individual scenarios.

For instance, the timing of the tipping point is subject to major shifts. Examining the drivers for the individual shifts leads to the conclusion that particularly the development of CO_2 prices exerts a decisive influence by increasing the costs of BF-BOF production. In contrast, H-DR cost reductions turned out insufficient for significantly increasing early H-DR attractiveness since the included hydrogen cost forecasts cause comparatively low approximations of H-DR to BF-BOF costs. Thus, at a general level, the increase in conventional steelmaking costs represents the essential prerequisite for achieving a cost advantage of hydrogen-based production.

[176] CP-B Scenario underlying.

[177] Cf. Fischedick et al., 2014; Toktarova et al., 2020.

[178] Cf. Facchini, Mossa, Mummolo, & Vitti, 2021.

Table 8: Sensitivity analysis of major diffusion factors.[176]

Variation / Variable	-8	-6	-4	-2	+2	+4	+6	+8
H-DR initiation	-34%	-26%	-18%	-9%	+9%	+18%	+26%	+33%
Diffusion speed	-18%	-13%	-9%	-4%	+5%	+9%	+13%	+17%

Color-coded values: change in total emissions caused by variation.

These correlations are confirmed when considering the underlying production costs of the scenarios. It was found that circumstances resulting in an early tipping point and thus an early initiation of H-DR production are simultaneously associated with increased production costs. However, it also became evident that these cost disadvantages are of temporary nature. By stimulating H-DR production, framework conditions that at first glance seem to negatively affect production costs might unlock additional cost reduction potential in the long term. This effect is mainly driven by the continuous reduction of hydrogen costs, accompanied by increasing BF-BOF costs, even occurring at low CO_2 prices, and a late phase-out of free allowances.

Nevertheless, these cost benefits of H-DR production are considerably smaller when BF-BOF production is not subject to any CO_2 allowance costs. If foreign producers are not obliged to purchase such allowances, German producers will face cost disadvantages in each scenario. Furthermore, it becomes apparent that future H-DR cost reductions can only compensate for these additional costs to a limited extent, thus indicating a long-lasting manifestation of these disadvantages. These findings reinforce the concerns about carbon leakage effects that might result in losses for the German steel industry. From a cost perspective, the influence of CO_2 prices must therefore be differentiated: On the one hand, a rapid rise and the associated increase in steel production costs lead to the earlier implementation of H-DR production. On the other hand, long-term cost disadvantages towards foreign producers would result.

A closer look at the insights of the emission forecasts reveals that the annual emission budgets are highly unlikely to be realized in the near future. Even the BC Scenario, subject to optimal conditions for an H-DR implementation, foresees compliance only from 2030 onwards. As a result, primary steel production turned out to be highly vulnerable to potential capacity reductions in case of strict compliance with the annual emission budget. Furthermore, the compensation potential of secondary production proved insufficient. In no case was the expansion of secondary production able to offset the required reduction in primary production to prevent the overall output from shrinking. A contradiction emerges from these findings: Strict enforcement of annual emission targets is incompatible with maintaining current production levels and requires one of the targets to be abandoned.

Looking beyond the exceedance of annual emissions budgets, a comprehensive transformation of primary production promises excellent opportunities to unleash enormous emissions savings and bring German steel production on track with its long-term emission targets. This potential is evident in the BC and EPO Scenarios, which both include achieving the long-term goal of carbon neutrality in 2045 despite temporarily exceeding emission budgets. The opposite occurs if current primary production is continued, as represented in the CP-D Scenario. BF-BOF production does not offer sufficient emission savings potential without reducing the production level, and thus proves to be the biggest obstacle to achieving the emission targets. Thus, the early reduction of BF-BOF capacities is identified as a prerequisite to aligning steel production with the German climate targets.

In summary, a positive picture emerges when assessing the potential of H-DR diffusion. In principle, a shift to H-DR production offers the prospect of achieving the long-term emission targets while maintaining the production volume of the German steel industry. Furthermore, such transformation raises the possibility of reaching a more favorable cost path, on which cost advantages compared to BF-BOF production could grow in the long term. However, the findings confirm the high risk of carbon leakage effects, which should not be underestimated. Exposing the steel industry to CO_2 prices would be associated with long-lasting cost disadvantages for German steel production.

4.4.3. Assessment of the Current Policy Framework

The previous section concludes that raising the cost of conventional production provides the most effective lever for increasing H-DR attractiveness at an early stage. Thus, the current policy framework is conceptually well suited to exert influence on steel production by regulating free allowances. Furthermore, the issuance of free allowances offers effective protection against cost disadvantages, as can be seen in the enormous effects of the phase-outs on production costs within the scenarios. However, potential future challenges associated with this regulation are apparent.

Since the EU ETS Regulation anticipates that the allocation of free allowances will be continuously reduced, this instrument provides only temporary support for involved emitters. In other sectors, such as aviation, loosening of the regulation has already been implemented, which suggests that steel production will also be affected at some point.[179] Thus, for German steel producers, it can be concluded that the current policy framework most likely only provides temporary protection in any case and merely delays the point in time

[179] Cf. ICAP, 2021, pp. 3-5.

when the disadvantage against international competitors is initiated. While this might provide producers time to transform their production processes, the emerging incentives are not entirely clear. Within the purely cost-based evaluation underlying, a prolonged issuance of free allowances delays H-DR diffusion as the key leverage for its attractiveness is suspended. The resulting extension of BF-BOF production leads to significantly higher overruns of annual emissions budgets and increases the likelihood of BF-BOF lock-in due to long-term investments being made. Long-term cost disadvantages also seem likely, as the scenario analysis showed that early H-DR implementation could enable achieving a more favorable cost path in the long term.

Furthermore, it is apparent that the current policy framework conflicts with the more incentive-based support via CCfDs: The calculation of required compensation payments showed that the distribution of free allowances causes significantly higher payments. This incompatibility leads to the result that ending the issuance of free allowances would increase the efficiency of more targeted measures.

In summary, the distribution of free allowances basically meets its objective of protecting the German steel industry from cost disadvantages in international competition. However, various complications were identified within the underlying modeling. The arising key issue is that the current policy framework prolongs the economic viability of BF-BOF production and thus delays the implementation of H-DR production. As discussed in the general findings, this might create long-term cost disadvantages. Combining this with the fact that the distribution of free allowances is only temporary in any case, these disadvantages might constitute the major long term impact of the current regulation. Hence, effects counteracting the actual policy target would result. Moreover, the current policy framework proves problematic from a climate policy point of view: The longer BF-BOF production is maintained, the more the emission targets are exceeded, and the more severe the consequences for the steel industry would turn out, should the targets eventually be strictly enforced.

4.4.4. Recommended Actions for Policy Makers

It has become clear that a serious climate policy, which also aims to preserve the German steel industry, must promote the earliest possible switch in primary steelmaking. This option is the only way to meet the defined emission targets without significantly reducing the steel output. Hydrogen-based steel production turned out to offer great potential for uniting these objectives, which is why it should be part of this political endeavor.

The developed scenarios show that the range of possible developments is still extensive, which can be justified by enormous underlying uncertainties. Of the scenarios, only the BC Scenario can be classified as unrealistic since an immediate end to the free allowances regime, and the availability of imported green hydrogen from 2026 onwards are not foreseeable. Both the CP D and the CP-B Scenario involve gross violations of the emission targets and would likely result in major losses in German steel production. Especially the CP-D Scenario turns out to be entirely incompatible with climate policy aspirations, which is why its materialization should be utterly prevented. On the contrary, the EPO Scenario would be associated with many desirable developments. It meets annual emission targets from 2035 onwards and enables realizing the long-term goal of carbon neutrality by 2045 while fully maintaining German steel output. Under currently foreseeable developments, reaching the EPO Scenario can thus be considered to be a realistic target.

Key parameters that should be focused on from a political perspective can be derived from the scenario assumptions. On a general level, the earliest possible promotion of adequate import infrastructure is a fundamental prerequisite, without which extensive H-DR production in Germany would probably be impossible. Subsidizing imported hydrogen would also directly impact H-DR costs and increase its attractiveness.

However, to promote H-DR diffusion as effectively as possible, the main focus of political efforts should lie in the role of CO_2 prices. Their development plays an essential role, whereby the step from the Stated Policies (STEPS) to the Announced Pledges (APS) scenario of the IEA proved to be critical. Based on this finding, recommendations can be derived from the applied IEA scenarios. Nevertheless, these recommendations can only be made at the level of the European Union, apart from steel industry-specific aspects, as decisions taken there primarily influence the EU ETS pricing. To close the gap between STEPS and APS forecasts, the IEA describes the current measures of the European Union as insufficient and recommends the full implementation of the proposed Fit-for-55 package.[180] Besides a stronger Emission Trading System and establishing new infrastructure for alternative energy carriers, this package includes numerous other measures.[181] Thus, it illustrates the need for a high degree of climate policy commitment at the European level as an essential factor for the future of German steel production.

The second lever that policymakers can use to regulate the effect of CO_2 prices is the distribution of free allowances. The underlying analysis has clearly shown that exposing steel production to the EU ETS is highly effective in increasing H-DR attractiveness at an early stage. Therefore, the early phase-out of free allowances described in the Fit for 55 package, as it is also subject to the EPO Scenario, would be beneficial. Additionally, such phase-out would allow efficient subsidization by other measures such as CCfDs, enabling a targeted and success-oriented promotion of H-DR production. As the findings suggest that opening steel production to the EU ETS would likely lead to increased production costs, the protection currently provided by free allowances must be replaced by other measures to counteract carbon leakage effects. For this purpose, a Carbon Border Adjustment Mechanism is a reasonable solution, as it would allow exposure to

[180] Cf. IEA, 2021, p. 170.
[181] Cf. EC, 2021c, p. 3.

the market mechanisms of the EU ETS while providing protection against international competition.

Lastly, the scenario analysis illustrates that focusing on long-term targets instead of annual emission budgets is most reasonable, as a temporary overshooting of emission budgets seems unavoidable and distracts from long-term developments. Instead, it becomes clear that once the transition to H-DR production is initiated, it offers excellent potential to align the industry with the long-term emission targets within a few years.

Summarizing the results, it emerges that the current policy framework can only temporarily fulfill its objective of protecting the costs of German steel from international competition. It does not provide distinct incentives for initiating a transformation of steel production and could even result in long-term disadvantages by delaying it. For this reason, it is recommended to adopt the initiation of H-DR production in Germany as the core policy objective, as the method offers great potential to achieve Germany's long-term emission targets while preserving the current output levels. Nevertheless, H-DR production needs extensive policy support to develop sufficient competitiveness with BF-BOF production. A vital prerequisite is establishing hydrogen import infrastructure to enable cost-competitive and large-scale H-DR steel production in Germany. However, the most effective lever for influencing the attractiveness of H-DR production proves to be the effects of CO_2 pricing on conventional steel. In addition to efforts at the European level to stimulate the increase in CO_2 prices through new climate policy measures, it is particularly recommended to terminate the free allowances regime for the steel industry. Such measures would significantly increase H-DR attractiveness and impose direct cost incentives on steel producers to abandon BF-BOF production. Furthermore, these would allow an effective application of other targeted support measures. Nevertheless, it must be acknowledged that exposing the steel industry to the EU ETS would result in significant cost disadvantages for producers. Thus, adequate measures for protection in the international market are required.

5. Conclusion and Outlook

The first research question focuses on identifying decisive drivers for the future role of hydrogen-based steelmaking in Germany. The first part of this thesis was devoted to answering this question. For this purpose, a qualitative analysis of the German steel industry's environment was carried out, through which technological, industry-internal, political, and other cost-influencing drivers were identified.

These results served to answer the second research question, which addresses the development of explorative scenarios for the diffusion of H-DR production in Germany. Within the modeling, four scenarios were extracted based on different combinations of the identified drivers: the Current Policy - Downside (CP-D), the Current Policy - Baseline (CP-B), the Early Phase Out (EPO), and the Best Case (BC) Scenario.

The analyzed scenarios differ considerably regarding observed model outputs, indicating that the future role of H-DR production is subject to significant uncertainties. While the CP-D Scenario does not provide sufficient incentives for H-DR implementation over the entire observation period, the other scenarios anticipate a comprehensive H-DR diffusion: Within the CP-B Scenario, H-DR production is implemented between 2035 and 2050, in the EPO scenario between 2030 and 2040, and in the BC Scenario between 2026 and 2036. The associated emission developments are also subject to strong deviations. The CP-D Scenario results in a massive overrun of annual emission budgets and proves to be completely incompatible with all emission targets. The CP-B Scenario likewise exceeds the annual emissions budgets until 2049 but reaches the goal of climate neutrality in 2050, five years later than required by current policy targets. The EPO and BC scenarios project annual emissions budgets to be undercut as of 2035 and 2030, thus both achieving the long-term goal of carbon neutrality by 2045. The development of CO_2 prices and the exposure of steel production to these were identified as the most effective levers for early H-DR promotion.

Major challenges arise from these findings. For instance, it seems unrealistic to achieve short-term emission targets without reducing production volumes. Furthermore, exposing steel production to CO_2 prices leads to increased production costs, suggesting disadvantages compared to producers that are not subject to this regulation.

To cope with these challenges, implications for policymaking were investigated, as was the aim of the third research question. Key recommendations are to focus on achieving long-term emission targets and stimulating the H-DR transformation. Only such transformation holds out the prospect of sufficient emission reductions while preserving the current level of industry output. To provide steel producers with distinct incentives for transformation and to enable the establishment of targeted policy measures, a shift from the current policy framework towards the earliest possible end of the free allowances regime is recommended. However, resulting cost disadvantages in the international market and the associated risk of carbon leakage effects must also be acknowledged. Thus, suitable mechanisms for its prevention must be established simultaneously.

Future research should identify and investigate the associated implications of concrete support measures such as CCfDs or Carbon Border Adjustment Mechanisms to better understand their suitability. For instance, an interesting approach could be determining the distribution of the incurred costs among the different actors and the resulting consequences. In addition, analyzing the market potential of green steel should be a core subject of future research. Practical approaches could be discussed in terms of the extent to which consumers might show an increased willingness to pay for green steel, how large the resulting markets might become, and how the creation of such markets could be promoted.

References

AGEB. (2021). *Stromerzeugung nach Energieträgern (Strommix) von 1990 bis 2021 (in TWh) Deutschland insgesamt: Datenstand Dezember 2021*. https://ag-energiebilanzen.de/wp-content/uploads/2021/02/Strommix-Dezember2021.pdf. (Accessed: 15.04.2022)

Agora Energiewende und Wuppertal Institut. (2019). *Klimaneutrale Industrie: Schlüsseltechnologien und Politikoptionen für Stahl, Chemie und Zement*. Berlin.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures*, *46*, 23–40.

ArcelorMittal. (2021). *Annual report 2020*. Luxembourg.

Arens, M., Åhman, M., & Vogl, V. (2021). Which countries are prepared to green their coal-based steel industry with electricity? - Reviewing climate and energy policy as well as the implementation of renewable electricity. *Renewable and Sustainable Energy Reviews*, *143*, 110938.

Arens, M., & Worrell, E. (2014). Diffusion of energy efficient technologies in the German steel industry and their impact on energy consumption. *Energy*, *73*, 968–977.

Arens, M., Worrell, E., Eichhammer, W., Hasanbeigi, A., & Zhang, Q. (2017). Pathways to a low-carbon iron and steel industry in the medium-term – the case of Germany. *Journal of Cleaner Production*, *163*, 84–98.

Bailera, M., Lisbona, P., Peña, B., & Romeo, L. M. (2021). A review on CO2 mitigation in the Iron and Steel industry through Power to X processes. *Journal of CO2 Utilization*, *46*, 101456.

Baracchini, G., Bianco, L., Cirilli, F., Echterhof, T., Griessacher, T., Millan-Agorio, M., … Sommerauer, H. (2019). *Biochar for a sustainable EAF steel production (GREENEAF2): Final report*. Directorate-General for Research and Innovation of the European Commission, Luxembourg.

Bea, F. X., & Haas, J. (2017). *Strategisches Management: XL-Ausgabe* (9th ed.). Tübingen.

Bhaskar, A., Assadi, M., & Nikpey Somehsaraei, H. (2020). Decarbonization of the iron and steel industry with direct reduction of iron ore with green hydrogen. *Energies*, *13*(3), 758.

Birat, J.-P. (2020). Society, materials, and the environment: The case of steel. *Metals*, *10*(3), 331.

BMWI. (2020a). *For a strong steel industry in Germany and Europe: The Steel Action Concept*. Federal Ministry for Economic Affairs and Energy, Berlin.

BMWI. (2020b). *The National Hydrogen Strategy*. Federal Ministry for Economic Affairs and Energy, Berlin.

BMWI. (2021a). *Bericht der Bundesregierung zur Umsetzung der Nationalen Wasserstoffstrategie, Bundesministerium für Wirtschaft und Energie*. https://www.bmwi.de/Redaktion/DE/Publikationen/Energie/bericht-der-bundesregierung-zur-umsetzung-der-nationalen-wasserstoffstrategie.pdf?__blob=publicationFile&v=16. (Accessed: 27.04.2022)

BMWI. (2021b). *Transformation der Stahlindustrie und Handlungskonzept Stahl, Bundesministerium für Wirtschaft und Energie*. https://www.bmwi.de/Redaktion/DE/Downloads/S-T/20210503-transformation-der-stahlindustrie-und-handlunskonzept-stahl.html. (Accessed: 27.04.2022)

Brändle, G., Schönfisch, M., & Schulte, S. (2020). Estimating long-term global supply costs for low-carbon hydrogen. *EWI Working Paper, No. 20/04*.

Branger, F., Quirion, P., & Chevallier, J. (2016). Carbon leakage and competitiveness of cement and steel industries under the EU ETS: Much ado about nothing. *The Energy Journal*, *37*(3), 109–135.

Carbon Market Watch. (2016). *Industry windfall profits from Europe's carbon market: How energy-intensive companies cashed in on their pollution at taxpayers' expense: Carbon Market Watch Policy Briefing*.

Coates, J. F. (2000). Scenario planning. *Technological Forecasting and Social Change*, *65*(1), 115–123.

Davies, S. W., & Diaz-Rainey, I. (2011). The patterns of induced diffusion: Evidence from the international diffusion of wind energy. *Technological Forecasting and Social Change*, *78*(7), 1227–1241.

DEHst. (2014-2021). *VET Reports 2013-2020, German Emissions Trading Authority*. https://www.dehst.de/EN/european-emissions-trading/installation-operators/2013-2020/2013-2020_reporting/VET_reports/vet-report_node.html;jsessionid=0309706BAF2C707B6A6BA8FF563C2114.2_cid331. (Accessed: 29.04.2022)

Demus, T., Reichel, T., Schulten, M., Echterhof, T., & Pfeifer, H. (2016). Increasing the sustainability of steel production in the electric arc furnace by substituting fossil coal with biochar agglomerates. *Ironmaking & Steelmaking*, *43*(8), 564–570.

Destatis. (2022). *Index der Großhandelsverkaufspreise nach Wirtschaftszweigen des Großhandels (WZ 2008): Lange Reihen der Fachserie 17, Reihe 6 von Januar 2005 bis März 2022*. Statistisches Bundesamt.

EC. (2019a). Commission delegated decision (EU) 2019/708: Supplementing Directive 2003/87/EC of the European Parliament and of the Council concerning the determination of sectors and subsectors deemed at risk of carbon leakage for the period 2021 to 2030. *Official Journal of the European Union*, *62*(L120), 20–26.

EC. (2019b). *The European Green Deal: COM(2019) 640 final*. European Commission, Brussels.

EC. (2020). *A hydrogen strategy for a climate-neutral Europe: COM(2020) 301 final*. European Commission, Brussels.

EC. (2021a). *Carbon border adjustment mechanism: Questions and answers, European Commission*. https://ec.europa.eu/commission/presscorner/detail/en/qanda_21_3661. (Accessed: 28.04.2022)

EC. (2021b). Commission decision (EU) 2021/355: Concerning national implementation measures for the transitional free allocation of greenhouse gas emission allowances in accordance with Article 11(3) of Directive 2003/87/EC of the European Parliament and of the Council. *Official Journal of the European Union*, *64*(L68), 221–226.

EC. (2021c). *'Fit for 55': Delivering the EU's 2030 climate target on the way to climate neutrality: COM(2021) 550*. European Commission, Brussels.

EC. (2021d). *Towards competitive and clean European steel: SWD(2021) 353 final*. European Commission, Brussels.

ECB. (2022). *Euro foreign exchange reference rates, European Central Bank*. https://www.ecb.europa.eu/stats/policy_and_exchange_rates/euro_reference_exchange_rates/html/eurofxref-graph-usd.en.html. (Accessed: 29.04.2022)

EEA. (2021). *EU Emissions Trading System (ETS) data viewer, European Environment Agency*. https://www.eea.europa.eu/data-and-maps/dashboards/emissions-trading-viewer-1. (Accessed: 28.04.2022)

Facchini, F., Mossa, G., Mummolo, G., & Vitti, M. (2021). An economic model to assess profitable scenarios of EAF-based steelmaking plants under uncertain conditions. *Energies*, *14*(21), 7395.

Fan, Z., & Friedmann, S. J. (2021). Low-carbon production of iron and steel: Technology options, economic assessment, and policy. *Joule*, *5*(4), 829–862.

Faurecia. (2021). *Faurecia will develop ultra-low CO2 seat structures in partnership with SSAB as its fossil-free steel supplier*. https://www.faurecia.com/en/newsroom/faurecia-will-develop-ultra-low-co2-seat-structures-partnership-ssab-its-fossil-free-steel-supplier. (Accessed: 28.04.2022)

Fidalgo, B., Berrueco, C., & Millan, M. (2015). Chars from agricultural wastes as greener fuels for electric arc furnaces. *Journal of Analytical and Applied Pyrolysis*, *113*, 274–280.

Fischedick, M., Marzinkowski, J., Winzer, P., & Weigel, M. (2014). Techno-economic evaluation of innovative steel production technologies. *Journal of Cleaner Production*, *84*, 563–580.

Germeshuizen, L. M., & Blom, P. W. E. (2013). A techno-economic evaluation of the use of hydrogen in a steel production process, utilizing nuclear process heat. *International Journal of Hydrogen Energy*, *38*(25), 10671–10682.

H2GS. (2022). *H2 Green Steel to build large-scale fossil-free steel plant in northern Sweden, H2 Green Steel*. https://www.h2greensteel.com/fossil-free-steel-plant. (Accessed: 28.04.2022)

Hertwich, E. G., Ali, S. H., Ciacci, L., Fishman, T., Heeren, N., Masanet, E. R., … Wolfram, P. (2019). Material efficiency strategies to reducing greenhouse gas emissions associated with buildings, vehicles, and electronics. *Environmental Research Letters*, *14*(4), 043004.

Hiltunen, E. (2009). Scenarios: Process and outcome. *Journal of Futures*

Studies, *13*(3), 151–152.
Holappa, L. (2020). A general vision for reduction of energy consumption and CO2 emissions from the steel industry. *Metals*, *10*(9), 1117.
HYBRIT. (2021). *SSAB, LKAB and Vattenfall to begin industrialization of future fossil-free steelmaking by establishing the world's first production plant for fossil-free sponge iron in Gällivare.* https://www.hybritdevelopment.se/en/march-24-2021-hybrit-ssab-lkab-and-vattenfall-to-begin-industrialization-of-future-fossil-free-steelmaking-by-establishing-the-worlds-first-production-plant-for-fossil-free-sponge-iron-in/. (Accessed: 28.04.2022)
ICAP. (2021). *ETS detailed information: EU Emissions Trading System (EU ETS), International Carbon Action Partnership.* https://icapcarbonaction.com/en/?option=com_etsmap&task=export&format=pdf&layout=list&systems%5B%5D=43. (Accessed: 15.03.2022)
ICAP. (2022). *Allowance price explorer, International Carbon Action Partnership.* https://icapcarbonaction.com/en/ets-prices. (Accessed: 15.03.2022)
IEA. (2019). *The future of hydrogen: Seizing today's opportunities, International Energy Agency*.
IEA. (2020). *Iron and steel technology roadmap: Towards more sustainable steelmaking, International Energy Agency*.
IEA. (2021). *World energy outlook 2021, International Energy Agency*.
IPCC. (2021). *Climate change 2021: The physical science basis: Contribution of working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change*.
Jacobasch, E., Herz, G., Rix, C., Müller, N., Reichelt, E., Jahn, M., & Michaelis, A. (2021). Economic evaluation of low-carbon steelmaking via coupling of electrolysis and direct reduction. *Journal of Cleaner Production*, *328*, 129502.
Karakaya, E., Nuur, C., & Assbring, L. (2018). Potential transitions in the iron and steel industry in Sweden: Towards a hydrogen-based future? *Journal of Cleaner Production*, *195*, 651–663.
Kirschen, M., Badr, K., & Pfeifer, H. (2011). Influence of direct reduced iron on the energy balance of the electric arc furnace in steel industry. *Energy*, *36*(10), 6146–6155.
Kushnir, D., Hansen, T., Vogl, V., & Åhman, M. (2020). Adopting hydrogen direct reduction for the Swedish steel industry: A technological innovation system (TIS) study. *Journal of Cleaner Production*, *242*, 118185.
Mandova, H., Patrizio, P., Leduc, S., Kjärstad, J., Wang, C., Wetterlund, E., … Gale, W. (2019). Achieving carbon-neutral iron and steelmaking in Europe through the deployment of bioenergy with carbon capture and storage. *Journal of Cleaner Production*, *218*, 118–129.
Mayer, J., Bachner, G., & Steininger, K. W. (2019). Macroeconomic implications of switching to process-emission-free iron and steel production in Europe. *Journal of Cleaner Production*, *210*, 1517–1533.
Medarac, H., Moya, J. A., & Somers, J. (2020). *Production costs from iron and steel industry in the EU and third countries, EUR 30316 EN*. Joint Research Centre of the European Commission, Luxembourg.
Mercedes-Benz Group. (2021). *From 2025: "Green" steel for Mercedes-Benz.* https://group.mercedes-benz.com/sustainability/climate/green-steel.html. (Accessed: 28.04.2022)
Miele. (2021). *Miele setzt auf grünen Stahl der Salzgitter AG.* https://www.miele.de/de/m/miele-setzt-auf-gruenen-stahl-der-salzgitter-ag-5751.htm? (Accessed: 28.04.2022)
Milford, R. L., Pauliuk, S., Allwood, J. M., & Müller, D. B. (2013). The roles of energy and material efficiency in meeting steel industry CO2 targets. *Environmental science & technology*, *47*(7), 3455–3462.
Morfeldt, J., Nijs, W., & Silveira, S. (2015). The impact of climate targets on future steel production – an analysis based on a global energy system model. *Journal of Cleaner Production*, *103*, 469–482.
Müller, N., Herz, G., Reichelt, E., Jahn, M., & Michaelis, A. (2021). Assessment of fossil-free steelmaking based on direct reduction applying high-temperature electrolysis. *Cleaner Engineering and Technology*, *4*, 100158.
Muslemani, H., Liang, X., Kaesehage, K., Ascui, F., & Wilson, J. (2021). Opportunities and challenges for decarbonizing steel production by creating markets for 'green steel' products. *Journal of Cleaner Production*, *315*, 128127.
Nilsson, L. J., Bauer, F., Åhman, M., Andersson, F. N. G., Bataille, C., de La Rue Can, S., … Vogl, V. (2021). An industrial policy framework for transforming energy and emissions intensive industries towards zero emissions. *Climate Policy*, *21*(8), 1053–1065.
Okereke, C., & McDaniels, D. (2012). To what extent are EU steel companies susceptible to competitive loss due to climate policy? *Energy Policy*, *46*, 203–215.
Otto, A., Robinius, M., Grube, T., Schiebahn, S., Praktiknjo, A., & Stolten, D. (2017). Power-to-Steel: Reducing CO2 through the integration of renewable energy and hydrogen into the German steel industry. *Energies*, *10*(4), 451.
Paltsev, S., Morris, J., Kheshgi, H., & Herzog, H. (2021). Hard-to-Abate sectors: The role of industrial carbon capture and storage (CCS) in emission mitigation. *Applied Energy*, *300*, 117322.
Patisson, F., & Mirgaux, O. (2020). Hydrogen ironmaking: How it works. *Metals*, *10*(7), 922.
Paul, J. (2015). *Praxisorientierte Einführung in die Allgemeine Betriebswirtschaftslehre: Mit Beispielen und Fallstudien* (3rd ed.). Wiesbaden, Springer Gabler.
Pauliuk, S., & Heeren, N. (2021). Material efficiency and its contribution to climate change mitigation in Germany: A deep decarbonization scenario analysis until 2060. *Journal of Industrial Ecology*, *25*(2), 479–493.
Pauliuk, S., Milford, R. L., Müller, D. B., & Allwood, J. M. (2013). The steel scrap age. *Environmental science & technology*, *47*(7), 3448–3454.
Pei, M., Petäjäniemi, M., Regnell, A., & Wijk, O. (2020). Toward a fossil free future with HYBRIT: Development of iron and steelmaking technology in Sweden and Finland. *Metals*, *10*(7), 972.
Richstein, J. C. (2017). *Project-based carbon contracts: A way to finance innovative low-carbon investments: Discussion Paper No. 1714*. German Institute for Economic Research.
Rogers, E. M. (1983). *Diffusion of innovations* (3rd ed.). New York.
Rootzén, J., & Johnsson, F. (2016). Paying the full price of steel – Perspectives on the cost of reducing carbon dioxide emissions from the steel industry. *Energy Policy*, *98*, 459–469.
Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity analysis in practice: A guide to assessing scientific models*. Chichester.
Salzgitter AG. (2021). *Annual report 2020*. Salzgitter.
Salzgitter AG. (2022a). *As from 2026, Salzgitter AG to deliver low-CO2 steel to all BMW Group plants in Europe and thus making a major contribution to supporting its customer in achieving their climate targets.* https://www.salzgitter-ag.com/en/newsroom/press-releases/details/translate-to-englisch-salzgitter-ag-liefert-ab-2026-co2-armen-stahl-an-alle-bmw-group-werke-in-europa-und-unterstuetzt-damit-ihren-kunden-massgeblich-beim-erreichen-seiner-klimaziele-19179.html. (Accessed: 28.04.2022)
Salzgitter AG. (2022b). *Volkswagen AG und Salzgitter AG vereinbaren die Lieferung von CO2-armem Stahl ab Ende 2025.* https://www.salzgitter-ag.com/de/newsroom/pressemeldungen/details/default-b8dd282ba9-19456.html. (Accessed: 03.05.2022)
Sartor, O., & Bataille, C. (2019). *Decarbonizing basic materials in Europe: How Carbon Contracts-for-Difference could help bring breakthrough technologies to market: Study No. 06/19*. Institute for Sustainable Development and International Relations.
Scania. (2021). *Fossil-free steel agiant step in Scania's decarbonisation.* https://www.scania.com/group/en/home/newsroom/news/2021/fossil-free-steel-a-giant-step-in-scanias-decarbonisation.html. (Accessed: 28.04.2022)
Schaeffler. (2021). *Schaeffler to buy green steel from Swedish start-up company H2greensteel.* https://www.schaeffler.com/content.schaeffler.com/en/news_media/press_releases/press_releases_detail.jsp?id=87743367. (Accessed: 28.04.2022)
Schoemaker, P. J. H. (1995). Scenario planning: A tool for strategic thinking. *Sloan Management Review*, *36*(2), 25–40.
Sidorov, S., Faizliev, A., Balash, V., Balash, O., Krylova, M., & Fomenko, A. (2021). Extended innovation diffusion models and their empirical performance on real propagation data. *Journal of Marketing Analytics*, *9*(2), 99–110.

SSAB. (2022). *SSAB joins forces with Polestar to develop a climate-neutral car*. https://www.ssab.com/news/2022/02/ssab-joins-forces-with-polestar-to-develop-a-climateneutral-car. (Accessed: 28.04.2022)

Sundt, S., & Rehdanz, K. (2014). *Consumer's willingness to pay for green electricity: A meta-analysis of the literature: Kiel Working Paper, No. 1931*. Kiel Institute for the World Economy.

Thyssenkrupp. (2021). *Annual report 2020/2021*. Essen.

Thyssenkrupp. (2022). *Into the future with green steel.* https://www.thyssenkrupp-steel.com/en/company/sustainability/climate-strategy/. (Accessed: 28.04.2022)

Toktarova, A., Göransson, L., & Johnsson, F. (2021). Design of clean steel production with hydrogen: Impact of electricity system composition. *Energies*, *14*(24), 8349.

Toktarova, A., Karlsson, I., Rootzén, J., Göransson, L., Odenberger, M., & Johnsson, F. (2020). Pathways for low-carbon transition of the steel industry - A Swedish case study. *Energies*, *13*(15), 3840.

UBA. (2021). *Entwicklung der spezifischen Kohlendioxid-Emissionen des deutschen Strommix in den Jahren 1990 - 2019: Climate Change 45/2021*. Umweltbundesamt.

UBA. (2022). *National trend tables in the structure of the sectors of the German Climate Protection Act (KSG) 1990-2020, Version for EU-Submission: 12.01.2022, German Environment Agency.* https://www.umweltbundesamt.de/daten/klima/treibhausgas-emissionen-in-deutschland#nationale-und-europaische-klimaziele. (Accessed: 28.04.2022)

UN. (2015). *Paris Agreement to the United Nations Framework Convention on Climate Change*. United Nations.

Vogl, V., Åhman, M., & Nilsson, L. J. (2018). Assessment of hydrogen direct reduction for fossil-free steelmaking. *Journal of Cleaner Production*, *203*, 736–745.

Vogl, V., Åhman, M., & Nilsson, L. J. (2021). The making of green steel in the EU: A policy evaluation for the early commercialization phase. *Climate Policy*, *21*(1), 78–92.

Volvo Group. (2022). *Green steel collaboration – Volvo Group, SSAB and Ovako.* https://www.volvogroup.com/en/future-of-transportation/going-fossil-free/green-steel-collaboration.html#:~:text=The%20Green%20steel%20collaboration%20is,to%20power%20fuel%2Dcell%20vehicles. (Accessed: 28.04.2022)

Wang, A., van der Leun, K., Peters, D., & Buseman, M. (2020). *European Hydrogen Backbone: How a dedicated hydrogen infrastructure can be created*. European Hydrogen Backbone.

Weber, W., Kabst, R., & Baum, M. (2018). *Einführung in die Betriebswirtschaftslehre* (10th ed.). Wiesbaden, Springer Gabler.

Weigel, M., Fischedick, M., Marzinkowski, J., & Winzer, P. (2016). Multicriteria analysis of primary steelmaking technologies. *Journal of Cleaner Production*, *112*, 1064–1076.

Wesseling, J. H., Lechtenböhmer, S., Åhman, M., Nilsson, L. J., Worrell, E., & Coenen, L. (2017). The transition of energy intensive processing industries towards deep decarbonization: Characteristics and implications for future research. *Renewable and Sustainable Energy Reviews*, *79*, 1303–1313.

Woertler, M., Schuler, F., Voigt, N., Schmidt, T., Dahlmann, P., Luengen, H. B., & Ghenda, J.-T. (2013). *Steel's contribution to a low-carbon Europe 2050: Technical and economic analysis of the sector's CO2 abatement potential*. The Boston Consulting Group and Steel Institute VDEh.

Worldsteel. (2009-2022). *Steel Statistical Yearbooks 2009-2021, World Steel Association.* https://www.worldsteel.org/steel-by-topic/statistics/steel-statistical-yearbook.html. (Accessed: 29.04.2022)

WV Stahl. (2021a). *Ein politischer Rahmen für die Dekarbonisierung der Stahlindustrie: Positionspapier, Wirtschaftsvereinigung Stahl.* https://www.stahl-online.de/wp-content/uploads/2021_06_22_Positionspapier_Rahmen-fuer-die-Transformation.pdf. (Accessed: 26.04.2022)

WV Stahl. (2021b). *Fakten zur Stahlindustrie in Deutschland 2021, Wirtschaftsvereinigung Stahl.* https://www.stahl-online.de/wp-content/uploads/WV-Stahl_Fakten-2021_RZ_Web_neu.pdf. (Accessed: 05.05.2022)

WV Stahl. (2021c). *Stahlschrott-Außenhandel, Statistischer Bericht 2021, Wirtschaftsvereinigung Stahl.* https://www.stahl-online.de/wp-content/uploads/2111-Statistischer-Bericht-Stahlschrott.pdf. (Accessed: 05.05.2022)

WV Stahl. (2022). *Rohstahlproduktion in Deutschland, Jahresbilanz 2021, Wirtschaftsvereinigung Stahl.* https://www.stahl-online.de/wp-content/uploads/2022_01_24_PM_Rohstahlerzeugung_Dez-21.pdf. (Accessed: 29.04.2022)

Wyns, T., Khandekar, G., Axelson, M., Sartor, O., & Neuhoff, K. (2019). *Industrial transformation 2050 - Towards an industrial strategy for a climate neutral Europe*. Institute for European Studies.

Xylia, M., Silveira, S., Duerinck, J., & Meinke-Hubeny, F. (2018). Weighing regional scrap availability in global pathways for steel production processes. *Energy Efficiency*, *11*(5), 1135–1159.

Junior Management Science 8(3) (2023) 717-731

Junior Management Science

journal homepage: www.jums.academy

The influence of pay transparency on organizational citizenship behavior

Felix M. Dietrich

Ludwig-Maximilians-Universität München

Abstract

Outcomes of recent institutional advances towards pay transparency on the individual level remain as unclear as resulting consequences for organizations due to scarce research. Particularly, the prevalent literature reveals a lack of studies which investigate the effect of pay transparency on organizational citizenship behavior (OCB). To provide clarity regarding the impact of pay transparency on OCB, this study uses a legislative change in Germany which requires the disclosure of selected pay-related information by regulated organizations to empirically investigate the transparency-OCB relation. To further address a common critique of pay transparency concerning potential negative effects of pay comparison among peers, relative standing is integrated as moderator to examine how the comparison of individuals' pay to that of referent others affects the link between pay transparency and OCB. Contrary to the theoretically derived expectations, pay transparency unfolds a significant negative effect on OCB while a moderation by relative standing cannot be supported. In joint consideration with other studies, findings imply that different forms of pay transparency can create varying consequences for OCB. Besides, the results aim to raise awareness among managers that disclosing pay-related information with little informativeness, or refrained disclosure are not optimal responses to regulatory changes towards more pay transparency as such strategies may result in a detrimental effect of pay transparency on OCB.

Keywords: Pay transparency; Organizational citizenship; Compensation.

1. Introduction

Despite late institutional pushes towards pay transparency in Europe (Veldman, 2017, p. 1) and North America (Belogolovsky & Bamberger, 2014, p. 1706) aiming to reduce socio-economic inequalities such as the gender pay gap, the actual effects of pay transparency on the individual level and resulting consequences for organizations are understudied (Colella, Paetzold, Zardkoohi, & Wesson, 2007, p. 55; Gupta & Shaw, 2014, p. 1; Marasi & Bennett, 2016, p. 50). Particularly, the link between pay transparency and organizational citizenship behavior (OCB) is ambiguous (SimanTov-Nachlieli & Bamberger, 2021, p. 230) although OCB constitutes a key component of job performance (Rotundo & Sackett, 2002, p. 66) and plays a vital role for organizational survival in the long term (Katz, 1964, p. 132). Regarding the literature, only three studies provide implications on the link between pay transparency and OCB. Marasi, Wall, and Bennett (2018, p. 70) infer from a statistically insignificant negative influence of pay secrecy on OCB that pay transparency creates a counter-directional, positive influence on OCB. Further, an explorative study by Göbel, Weller, and Nyberg (2020, p. 6) empirically demonstrates that pay transparency negatively affects OCB. Findings of the latter correspond to those of Bamberger and Belogolovsky (2017, p. 658) who report a detrimental impact of pay transparency on helping behavior which is a central component of OCB (Organ, 1988, p. 4). Beyond the discursive inconsistency, critics of pay transparency argue that transparent pay structures may reveal potential differences in wages among employees and thus can result in detrimental behavioral outcomes such as reduced cooperation (Colella et al., 2007, p. 55). Yet, research on individuals' standing regarding their pay relative to that of peers as moderator of the pay transparency-OCB relation has not been tested yet. Hence, the aim for progressive coherence while reducing incompleteness of the pay transparency literature academically motivates the research question: Does pay transparency relate to OCB and does relative standing influence the relationship between the two concepts?

To empirically clarify the relation between pay trans-

DOI: https://doi.org/10.5282/jums/v8i3pp717-731

parency and OCB, this research uses the introduction of a legislative change in Germany called "Transparency in Wage Structures Act" (TWSA) which requires regulated organizations to disclose selected pay-related information as pay transparency condition in a quasi-experiment. The effect of pay transparency on OCB is examined with a difference-in-differences estimation (DD). Subsequently, the moderation of the two concepts by relative standing of individuals' pay compared to that of peers is analyzed with a difference-in-difference-in-differences estimation (DDD). Contrary to expectations derived from social exchange (Blau, 1984, p. 1), social comparison (Festinger, 1954, p. 117), and attribution theory (Miller & Ross, 1975, p. 213), results show a significant negative effect of pay transparency on OCB while the moderating role of relative standing cannot be supported. Whereas prevalent studies focus on extreme forms of pay transparency (Brown, Nyberg, Weller, & Strizver, 2022, p. 10), this study adds to the pay transparency literature by empirically investigating the effect of selective pay transparency as created by the TWSA on OCB. Further, this paper proposes that the selective pay transparency instated by the TWSA may not provide individuals with sufficient information to induce a significant moderating role of relative standing. Accordingly, the common critique that pay comparison among peers stimulated by pay transparency evokes negative behavioral consequences may not be fundamentally applicable. Regarding the practical motivation of this study, the findings aim to provide managers with strategic implications how to mitigate the negative effect of selective pay transparency as generated by the TWSA on OCB amidst a current trend towards more transparent wage structures (Belogolovsky & Bamberger, 2014, p. 1706; Veldman, 2017, p. 1).

The following segment provides an overview on the theoretical background of the concepts used to investigate the effects of pay transparency on OCB before contextualizing relative standing as potential moderator. Subsequently, methodology and details of the empirical analyses are presented, followed by the results. Finally, discoveries are discussed, and a summarizing conclusion is drawn.

2. Theoretical background

2.1. Pay transparency

Pay transparency can be defined as an equilibrium state without information asymmetry among actors (e.g. employer and employees) who possess pay-related information (Göbel et al., 2020, p. 1). Hence, pay transparency provides employees with unrestricted knowledge about other organizational members' pay (Brown et al., 2022, p. 3).

Pay transparency is promoted by high pay information disclosure which refers to the act of communicating relevant pay-related information (Brown et al., 2022, p. 3). Fulmer and Chen (2014, p. 169), evaluate pay information disclosure according to the restrictiveness of the communicated information on the pay allocation (pay-outcome transparency), on the process determining the pay distribution (pay-process transparency), and the liberty to share pay-related information with others (pay-communication transparency). Examined more closely, the assessment of pay information disclosure along the three orthogonal dimensions proposed by Fulmer and Chen (2014, p. 169) depends on the quantity, quality, and timing of the disclosed pay-related information (Brown et al., 2022, p. 5). Quantity indicates the amount of pay-related information accessible to an actor. Whereas some organizations provide their employees with pay-related information about every organizational member, other companies disclose no pay-related information (Marasi & Bennett, 2016, p. 52). Quality refers to the specificity of the reported metrics. Hereof, the informational content of individuals' exact salary is richer than the one of aggregated metrics such as median pay (Colella et al., 2007, p. 58; Montag-Smit & Smit, 2021, p. 709; Smit & Montag-Smit, 2019, p. 538). Lastly, the timing dimension pertains the moment or period of disclosing pay information, e.g. during the recruiting of potential employees or during pay raise negotiations with existing employees. In summary, constant disclosure of high quantity and quality information on the pay allocation as well as on the underlying distribution process in combination with the freedom to exchange the obtained information with others inside and outside the organization pushes the receiver of pay-related information towards an informational state of pay transparency. Opposingly, persistent restrictive pay information disclosure characterized by low quantity and quality of pay-related information on the pay allocation as well as distributional process in interplay with a ban on exchanging the obtained information with others evokes a shift towards pay secrecy. Furthermore, studies conducted by Göbel, Weller, and Nyberg (n.d., p. 4) as well as Marasi and Bennett (2016, p. 52) investigated the multi-directionality of pay information disclosure. Both conclude that the sharing of pay-related information can occur through two distinct interactional channels, implying that employers and employees co-create pay transparency (Göbel et al., n.d., p. 4). On the one hand, a unidirectional exchange between the organization as sender and its employees as receivers of pay-related information promotes pay transparency in a top-down manner. On the other hand, employees exchange pay-related information multilaterally among each other and thus facilitate pay transparency emergently. In summary, sender-receiver interactions may be heterogenous regarding the participants of the information exchange and differ concerning the quality, quantity, and timing of the disclosed pay information. This can lead to different peculiarities on the three orthogonal dimensions of pay transparency proposed by Fulmer and Chen (2014, p. 169). Resultingly, a continuum ranging from pay transparency to pay secrecy is spanned with various incremental forms of the two extremes in between (Brown et al., 2022, pp. 3–4).

Assessing whether organizational policies and practices tend towards pay transparency or secrecy is relevant as both extremes and their incremental forms cause fundamentally different perceptions and attitudes. Policies and practices

related to pay secrecy induce employees to assume trust-reducing, malevolent intentions by their employer (Belogolovsky & Bamberger, 2014, p. 1708; Montag-Smit & Smit, 2021, p. 723). In line with the conceptualization of pay transparency and pay secrecy as opposites, scholars found that transparency concerning the pay allocation and the underlying distribution process enhances perceived fairness (Castilla, 2015, p. 328) and job satisfaction (Day, 2011, pp. 479–480; Futrell & Jenkins, 1978, p. 218) due to increased comprehensibility of pay-related decisions. Further, pay transparency facilitates trust because employees tend to interpret the disclosure of pay information by the employer as benevolent reduction of uncertainty concerning the link between performance and rewards (Belogolovsky & Bamberger, 2014, p. 1708; Montag-Smit & Smit, 2021, pp. 722–723; Schnackenberg & Tomlinson, 2016, p. 1797). However, Cullen and Perez-Truglia (2018, p. 39) empirically demonstrated that pay transparency generates perceived unfairness and reduces job satisfaction in case of unequal pay distributions among individuals in similar organizational roles because individuals tend to perceive the reasons of unequal peer pay distributions as non-meritocratic. Given these circumstances, a more secretive disclosure of pay-related information which does not allow inferences on the pay distribution among peers would avoid the detrimental effects described by Cullen and Perez-Truglia (2018, pp. 4–6). In sum, different forms of pay transparency and pay secrecy can create favorable as well as adverse effects.

2.2. Organizational citizenship behavior (OCB)

OCB describes voluntary extra-role behavior which positively impacts organizational effectiveness in sum but is not formally rewarded (Organ, 1988, p. 4). According to Smith, Organ, and Near (1983, p. 658), OCB can be directed at the organization (e.g. generalized compliance) and channeled towards individuals (e.g. support of coworkers). In addition to compliance and helping behavior, Organ (1988, p. 4) introduced courtesy, sportsmanship, and civic virtue as further dimensions of OCB. Courtesy refers to behavior which prevents cooperation and coordination problems (Konovsky & Organ, 1996, p. 255). Employees who handle work-related issues with positivity and resilience are characterized by sportsmanship (Konovsky & Organ, 1996, p. 255). Lastly, civic virtue relates to constructive participation in group or organizational issues (Konovsky & Organ, 1996, p. 257). Synopsized, the discretionary efforts associated with OCB go beyond formally required in-role performance for task completion.

The roots of OCB lie in role theory. Roles describe a bundle of behavioral expectations concerning a position within a social system (Dreitzel, 1980, p. 44; Nienhüser, 1993, p. 239). An organization regarded as social system rewards behavior displayed by individuals which is compliant with their jobs as social roles and punishes deviations from behavioral expectations (Dreitzel, 1980, p. 46; Matiaske, Wallmeier, & Weller, 2017, p. 256). Accordingly, an organization's sanctioning power induces employees to fulfill their job duties explicated in job descriptions and labor contracts with in-role behavior (Weller, Matiaske, & Holtmann, 2007, p. 176). As formal contracts are notoriously incomplete, extra-role behavior which describes efforts beyond formally required in-role behavior becomes attached to jobs and enables organizations to cope with unexpected challenges outside the scope of their employees' role prescriptions (Katz, 1964, p. 132; Weller et al., 2007, p. 176). Consequently, extra-role behavior is essential for organizational effectiveness because discretionary efforts facilitate organizational contingency adoption (Burns & Stalker, 2001, pp. 103–108) and therewith assists organizational survival in the long-term.

Among the different concepts under the headline of extra-role behavior, the notion of OCB introduced by Organ (1988, p. 4) constitutes probably the most prominent type of extra-role behavior (Matiaske et al., 2017, p. 263). Although OCB is formally not part of the reward system, high OCB is associated with a positive effect on formal performance ratings and consequently fosters pay rises as well as promotions (MacKenzie, Podsakoff, & Fetter, 1993, p. 76). In turn, Schnake and Dumler (1997, p. 222) demonstrated that the compensation enhances OCB. Building on the give and take notion implied by MacKenzie et al. (1993, p. 76) as well as Schnake and Dumler (1997, p. 222), OCB can be embedded in the context of social exchange theory (Blau, 1984, p. 1). Social exchange processes encourage extra-role behavior whereas economic exchange induces in-role behavior (Matiaske & Weller, 2007, p. 515; Organ, Podsakoff, & MacKenzie, 2006, pp. 54–55). Moreover, different media are utilized to conduct the two types of exchange. Labor contracts are used to explicate the components of the economic exchange including expected in-role behavior and compensation ex-ante to labor provision (e.g. monthly wages in arrears to labor provision). Accordingly, compensation prompts in-role behavior (Matiaske & Weller, 2007, p. 515). Contrastingly, the elements of social exchange are rooted in an implicit psychological contract with unspecified conditions ex-ante to contracting (Rousseau, 1995, pp. 23–54). Rather, diffuse expectations concerning future obligations emerge during relational interactions between actors such as employees (e.g. expecting fair compensation) and employers (e.g. expecting OCB). Although the social and economic channel are reciprocal, the party which provides advance concessions within the economic exchange expects short-term compensation as agreed a-priori to contracting. Conversely to the transactional character of the economic exchange, the social exchange builds on mutual trust in the return of an appropriate compensation to the discretion of the exchange partner in the longer-term (Organ et al., 2006, pp. 54–55). Further elaborating on the antecedents of OCB, meta-analyses by LePine, Erez, and Johnson (2002, p. 59) and Organ and Ryan (1995, p. 787) found fairness and job satisfaction to promote OCB as both concepts induce reciprocation by individuals. Also, individual studies conducted by Konovsky and Pugh (1994, p. 664) as well as Colquitt, LePine, Piccolo,

Zapata, and Rich (2012, pp. 4–5) found empirical evidence that trust facilitates OCB as amplifier of the social exchange between the individual and the organization.

2.3. Relative standing concerning peer pay

First experiments on pay perceptions under conditions of pay secrecy by Lawler (1965, pp. 417–419, 1967, pp. 187–188) indicate that employees overestimate the pay of equal or lower positions within the organizational hierarchy. Conversely, employees tend to underestimate the pay of higher positions. Building on Lawler's (1965, pp. 417–419, 1967, pp. 187–188) findings, Cullen and Perez-Truglia (2018, pp. 4–6) assessed the behavioral impact of employees' distorted pay perception. The authors found that an increase of 10% in perceived peer pay (i.e. the pay of employees in a similar position) would decrease work hours by 9.4%, given that the focal individual's pay remains unchanged. Critics who thematize negative effects of pay transparency argue that transparent wage structures reveal potentially unequal pay allocations among peers and thus create negative effects of peer pay comparison as described by Cullen and Perez-Truglia (2018, pp. 4–6) and multiple other studies on job satisfaction (Card, Mas, Moretti, & Saez, 2012, pp. 2995–2996), fairness perception and productivity (Breza, Kaur, & Shamdasani, 2018, pp. 624–627), as well as turnover (Dube, Giuliano, & Leonard, 2019, p. 639).

The socio-psychological motive behind peer pay comparison lies in the inherent drive for comparison with referent others (Festinger, 1954, pp. 117–118). Hereof, positional goods such as pay serve as medium for comparison among peers (Frank, 1985, p. 101). Driven by uncertainty aversiveness (van den Bos & Lind, 2002, pp. 6–7), the disclosed pay-related information induces individuals to engage in pay comparison among peers for a more accurate determination of their relative standing. Exploring perceptional reactions to pay transparency, SimanTov-Nachlieli and Bamberger (2021, p. 237) found that perceived distributive justice of the pay allocation is dependent on the relative standing of individuals. Paying employees one standard deviation unit less than the mean pay of their reference group resulted in a significant negative effect on perceived distributive justice (SimanTov-Nachlieli & Bamberger, 2021, p. 237). Contrastingly, paying employees one standard deviation unit more than the mean pay led to a positive, yet statistically insignificant effect on distributive justice (SimanTov-Nachlieli & Bamberger, 2021, p. 237). Hereof, Cullen and Perez-Truglia (2018, p. 39) suggest that the negative effect on distributive justice emerges because individuals with a deprived relative standing tend to perceive the reasons behind an unequal pay distribution among peers to be non-meritocratic (i.e. gender bias or favoritism) and thus unfair. The perceived lack of procedural justice can be explained with a cognition-based perspective. The self-attribution error posits that people tend to attribute detrimental outcomes to external factors outside their scope of control (Miller & Ross, 1975, pp. 213–214). This psychological coping mechanism allows individuals to explain their deprived standing while maintaining self-esteem (Heider, 1958, p. 173; Zuckerman, 1979, pp. 246–247) and perceived control over their environment (Kelley, 1971, p. 23; Langer & Roth, 1975, p. 951). Hence, the self-attribution error provides a possible explanation why employees tend to perceive the reason for unequal peer pay as non-meritocratic as supposed by Cullen and Perez-Truglia (2018, p. 39). The resulting perception of distributive unfairness leads employees with deprived relative standing to reduce their trust in an appropriate reward for their efforts (Austin, McGinn, & Susmilch, 1980, p. 439). In line with this theorizing, Brown, Ferris, Heller, and Keeping (2007, p. 67) provided empirical evidence of a negative association between upward peer pay comparison and job satisfaction. Regarding beneficial outcomes in contrast, the fundamental attribution error predicts that individuals attribute favorable occurrences to personal abilities due to overconfidence (Johnson & Fowler, 2011, p. 317; Miller & Ross, 1975, pp. 213–214). Accordingly, individuals with a beneficial relative standing tend to interpret the reasons behind their above-average pay to be meritocratic (i.e. based on performance and abilities), reinforcing the resulting perceived distributive justice of the pay allocation (Austin et al., 1980, p. 439). In analogy to this argumentation, SimanTov-Nachlieli and Bamberger (2021, p. 237) found that perceived distributive justice is higher among employees whose pay exceeds that of peers. Moreover, the feedback effect of reward positively reinforces trust among employees with relatively higher pay that future extraordinary efforts are recognized and rewarded appropriately by the organization (Matiaske & Weller, 2007, p. 516). Corresponding to this logic, Brown et al. (2007, p. 67) demonstrated that downward peer pay comparison enhances job satisfaction.

2.4. Effect of pay transparency on OCB

Following a social exchange perspective, compensation operates on the economic channel as extrinsic motivator for in-role performance (Matiaske & Weller, 2007, p. 515). Hence, Blau's (1984, p. 1) notion does not predict a positive influence of compensation on extra-role behavior as the latter is induced via processes on the social exchange channel. However, pay transparency adds a social dimension to compensation. Employees interpret the organization's motive behind the disclosure of pay related information as benevolent (Montag-Smit & Smit, 2021, pp. 722–723) because pay transparency raises the expectation that policies and practices determining the pay allocation (i.e. performance appraisals) are applied consistently and according to meritocracy (Castilla, 2015, p. 328). Accordingly, pay transparency reduces the uncertainty of the link between compensation and performance and thus improves trust in appropriate reward for future contributions (Belogolovsky & Bamberger, 2014, p. 1708; Schnackenberg & Tomlinson, 2016, p. 1797). The resulting perception of a fair pay allocation deploys a positive attitudinal effect on job satisfaction (Day, 2011, pp. 479–480; Futrell & Jenkins, 1978, p. 218). Accordingly, the employee's psychological contract with the organization (Rousseau, 1995, pp. 23–54) induces a feeling of unspecified

obligation to balance out the social exchange (Heider, 1958, p. 173) which may be fulfilled with reciprocating OCB since trust, perceived fairness, and job satisfaction antecede OCB (Colquitt et al., 2012, pp. 4–5; LePine et al., 2002, p. 59). Resultingly, pay transparency may create a positive effect on OCB:

> Hypothesis 1: Pay transparency positively affects OCB.

2.5. Relative standing as moderator

The increase of pay-related information coming along with pay transparency motivates social comparison among peers due to more accurate information on pay as positional good (Festinger, 1954, pp. 117–118; Frank, 1985, p. 101; van den Bos & Lind, 2002, pp. 6–7). Further, social processes induce the formation of fairness perceptions (Colquitt, Conlon, Wesson, Porter, & Ng, 2001, p. 426). Employees who earn more than their peers perceive that the organization fulfilled its part of the psychological contract regarding appropriate compensation as individuals attribute their beneficial relative standing to their merit performance (Miller & Ross, 1975, pp. 213–214). The resulting trust in appropriate compensation reinforces the perception of high distributive justice and facilitates job satisfaction (Day, 2011, pp. 479–480; Futrell & Jenkins, 1978, p. 218). To balance out the organization's contribution to the social exchange, employees who earn more than their peers may further increase their OCB:

> Hypothesis 2.1: The positive effect of pay transparency on OCB is stronger for individuals who earn more than their peers.

In contrast, employees who earn less than their peers experience a breach of the psychological contract as they perceive that the organization failed to compensate performance appropriately. Thereby, individuals tend to attribute their deprived relative standing to non-merit factors outside of their control (Miller & Ross, 1975, pp. 213–214). The resulting mistrust in appropriate compensation reduces perceived distributive justice and further job satisfaction (Day, 2011, pp. 479–480; Futrell & Jenkins, 1978, p. 218). Consequently, employees with deprived relative standing perceive their social exchange with the organization as unbalanced and thus may reduce their OCB to restore an equitable employer-employee relationship:

> Hypothesis 2.2: The positive effect of pay transparency on OCB is weaker or becomes negative for individuals who earn less than their peers.

3. Methodology and empirical analyses

3.1. Context of the data collection: The German Transparency in Wage Structures Act

Late pushes by institutional forces across Europe towards pay transparency (Veldman, 2017, p. 1) translated to the TWSA in Germany. The law was introduced in mid-2017 to reduce the gender pay gap (Bundesministerium für Familie, Senioren, Frauen und Jugend, 2017, p. 4). After a preliminary preparation time ending in early 2018, employees of German enterprises with a workforce greater than 200 people became equipped with the legal right to pro-actively request the median gross monthly pay and up to two further salary components such as benefits (Bundesministerium für Justiz, 2022, p. 1). The information is based on a reference group of at least six organizational members in a similar position but of the opposite-sex (Bundesministerium für Justiz, 2022, p. 1). Additionally, the employer must provide information about the process used for salary determination when an employee places a request based on the TWSA (Bundesministerium für Justiz, 2022, p. 1). Applying the typology by Brown et al. (2022, p. 5), the quality (mean pay), quantity (pay-related information restricted by gender and job similarity), and timing (pay-related information only requestable after entering an organization) of the disclosed pay information as required by the TWSA remain limited. Yet, the TWSA provides increased pay outcome as well as pay process transparency and has an indirect positive effect on pay communication transparency as the legislative change has been found to increase an emergent exchange of pay-related information among employees (Fulmer & Chen, 2014, p. 169; Göbel et al., n.d., p. 4). Taking the specifications of the legally required pay information disclosure into account, the TWSA creates a form of selective pay transparency.

3.2. Sample

This study is based on independent, cross-sectional survey data on the individual level gathered with an online survey described further in detail by Göbel et al. (n.d., pp. 14–15). The collected data is randomly drawn to reflect the average situation of German employees at the time of the data collection in five waves from December 2017 to 2020. The survey was emitted to employees of organizations affected by the TWSA (> 200 employees as treatment group) and employees of companies outside of the TWSA's scope (<= 200 as control group) while targeting a balanced ratio of treatment ($n_{treated} = 2510$) to control ($n_{control} = 2518$) group observations. The size of employers was limited to a range of 50 to 500. Participants who did not indicate yearly gross salary, size of their employer, industry, education, location, or OCB-related items were excluded from empirical analyses. The final sample used for analysis pools independent cross-sections and consists of $n = 5028$ observations.

3.3. Estimation strategy

This study uses the TWSA as shock in a quasi-experiment which introduces a pay transparency condition to regulated firms whereas companies outside of the law's scope remain unaffected. To examine the impact of pay transparency on OCB as perceived by individuals, this study deploys a difference-in-differences (DD) estimation based on multivariate linear regressions using ordinary-least squares (OLS)

for estimation (Wooldridge, 2019, pp. 431–436) to compare the development of mean OCB from before to after the introduction of the TWSA between treatment and control group. Consequently, inferences about the effect of the TWSA as transparency condition on OCB can be drawn. Results are subsequently utilized to support or reject hypothesis 1. To test hypotheses 2 and 3, a difference-in-difference-in-differences estimator (DDD) also based on multivariate linear regression using OLS for estimation (Wooldridge, 2019, pp. 436–437) is applied to support or reject the moderation of the pay transparency-OCB link by relative standing.

3.4. Dependent variables

The measurement of OCB as perceived by respondents was conducted with six items (Lee & Allen, 2002, p. 131). Answers ranged from 1 "Don't' agree" to 5 "fully agree" on a five-point scale. Half of the OCB items measured OCB aimed at individuals whereas the other half measured OCB directed at the organization. For analyses, the OCB subscales were aggregated to one OCB metric for each respondent by computing the mean over the single OCB items.

3.5. Independent variables

3.5.1. Treatment effect and moderation

The model used to test hypothesis 1 concerning the effect of pay transparency on OCB introduces the treatment effect $Time \times Treated$ which contains the interaction of the treatment group with the post-treatment period. The treatment group dummy $Treated$ becomes 0 when individuals indicated that their organization employs 200 or less employees and thus is not regulated by the TWSA. Contrastingly, the $Treated$ dummy takes the value of 1 when the organization employs 201 or more workers and hence is affected by the TWSA. The post-treatment dummy $Time$ corresponds to 0 during the period prior to the introduction of the TWSA in year 2017 and is labelled 1 in the period after the treatment from years 2018 to 2020.

Hypotheses 2 and 3 referring to a potential moderating effect of relative standing are tested with the DDD model which additionally introduces $Relative\ standing$ as variable to examine a triple interaction of $Time \times Relative\ standing \times Treated$. $Relative\ standing$ constitutes a dummy variable which corresponds to 1 when employees earn less than their reference pay and takes the value of 0 otherwise. The underlying reference pay was computed in two ways. Models 2.1 - 2.3 (Table 4) use the median pay of the sample to assign values to the $Relative\ standing$ dummy. Models 3.1 - 3.3 (Table 4) follow research by Schmidt (2017, pp. 10–13) to compute the reference pay based on comparable individuals.[1] Hereof, the reference pay for individuals is determined using the residuals of a multivariate linear regression to estimate the yearly gross pay[2] with OLS dependent on demographic differences (age[3], location[4]), human capital variables (education[5], leadership role[6], position[7]), and industry differences (size of the employer[8], industry sector[9]). Positive residuals identify employees who earn less than the expected average pay of individuals with comparable characteristics. In this case, the dummy $Relative\ standing$ takes the value of 1. Contrastingly, negative residuals depict employees who earn a salary higher than or equal to the expected average salary of individuals with similar characteristics. In such cases, the dummy $Relative\ standing$ takes the value of 0.

3.5.2. Control variables

To demonstrate the robustness of the estimated effects, three different model versions with no (Models 1.1, 2.1, 3.1), partial (Models 1.2, 2.2, 3.2), and full (Models 1.3, 2.3, 3.3) control sets are included in Tables 3 and 4. Interpretations of the empirical results are based on the models with partial controls which account for age as maturing may influence altruism and thus OCB (Wagner & Rush, 2000, p. 379). Further, the models control for the existence of a human resource department[10] because human resource professionals may actively design policies and practices to influence OCB and thus dilute the effect of the TWSA. Moreover, the covariate collective agreement[11] is included since employees with salaries determined by a collective agreement already know the salaries of their colleagues prior to the TWSA. This may distort the effect of pay transparency as created by the TWSA on OCB. Additionally, leadership roles are controlled for as this investigation focuses on horizontal peer pay comparison instead of vertical pay comparison to managers. The variable industry is introduced to extract varying levels of OCB across business sectors (Podsakoff, MacKenzie, Paine, & Bachrach, 2000, p. 513). Wave[11] controls for time fixed effects. In accordance with Göbel et al. (n.d., pp. 16–17), the models with full control variables additionally contain education, location, size of the employer, works council[12], union membership[13], sex[14], knowledge of the TWSA[15], knowledge of the term gender pay gap[16], as well as requests[17] to account for systematic differences in the treatment and control group

[1] The multivariate linear regression used to determine individuals' reference pay can be replicated using the code appended separately to this study.

[2] Measured in €

[3] Measured in years

[4] 0 = East Germany and 1 = West Germany

[5] 1 = no degree, 2 = early high-school drop-out, 3 = secondary school degree, 4 = college/university qualification, and 5 = academic degree

[6] 0 = no and 1 = yes

[7] 1 = apprentice, 2 = intern, 3 = trainee, junior, 4 = Trained employee without completed vocational training, 5 = Specialist with at least two years of completed training, 6 = Third management level e.g. team leader, 7 = Second management level e.g. head of department, 8 = First management level e.g. CEO, 9 = other

[8] Measured by headcount

[9] 1 = agriculture, forestry, and fishing, 2 = mining, 3 = energy and water supply, 4 = manufacturing industry, 5 = construction industry, 6 = trade, 7 = traffic, 8 = information and communication, 9 = tourism, accommodation and gastronomy, 10 = finance and Insurance, 11 = economical, scientific and self-employed services, 12 = education, 13 = health and social services, 14 = other services, 15 = public administration, 16 = other

[10], 12-17 0 = no and 1 = yes

[11] 1–5

and to further to reduce error variance in order to improve the precision of estimates.

4. Results

Table 1 reports descriptive statistics and Table 2 depicts correlations for variables used in the DD and DDD estimation models. Results of the econometric models used for hypotheses testing are reported in Tables 3 and 4.

Visualizing the data provides first insights to the effect of pay transparency on OCB as posited in hypothesis 1. Figure 1 shows the development of mean OCB over time, pooled according to treatment and control groups, and surrounded by error bands indicating the 90% confidence interval. The treatment group's drop in OCB occurs in timely conjunction with the introduction of the TWSA in 2018. The mean OCB of the treatment group does not recover to its pre-2018 level and seems to be persistently lower than the mean OCB in the control group after the TWSA became effective in 2018. Contrastingly, mean OCB of the control group remains seemingly unaffected in 2018. To further investigate the relationship between pay transparency and OCB, algebraic modelling and the results of multivariate linear regression analyses are introduced in the next paragraph.

$\widehat{\delta}_1$ depicts the DD estimator which allows to test hypothesis 1 by estimating the average treatment effect of pay transparency on OCB

$$\begin{aligned}\widehat{\delta}_1 &= \left(\overline{OCB}_{Time=1,Treated=1} - \overline{OCB}_{Time=0,Treated=1}\right) \\ &\quad - \left(\overline{OCB}_{Time=1,Treated=0} - \overline{OCB}_{Time=0,Treated=0}\right) \\ &= \widehat{\delta}_{DD}. \end{aligned} \tag{1}$$

To obtain an estimate for the causal impact of pay transparency on OCB, average OCB is computed before and after the introduction of the TWSA for organizations with more than 200 employees and organizations with 200 or less employees and subsequently subtracted. The result of $\widehat{\delta}_1$ can be obtained by pooling the data over the dummies $Time$ and $Treated$ and running a multivariate linear regression using OLS to estimate[12]

$$\begin{aligned}OCB_{it} = \beta_0 + \delta_0 Time_t + \beta_1 Treated_i + \delta_1 Time \\ \times Treated_{it} + Controls_{it} + u_{it}. \end{aligned} \tag{2}$$

The intercept β_0 represents the average OCB of employees in organizations with 200 or less workers before the TWSA was introduced. The parameter δ_0 captures the change in OCB of all employees in the sample from the pre-TWSA to the post-TWSA period to isolate the development of OCB over time which is not caused by the TWSA. β_1 measures the difference in OCB between organizations with 201 or more and 200 or less workers that is not due to the TWSA. Primary focus of the analysis is the parameter $\delta_1 Time \times Treated$ which measures the change in OCB due to the TWSA, assuming that OCB did not systematically change for other reasons than the ones controlled for in Models 1.2 and 1.3 (Table 3). To examine the effect size, direction, and statistical significance of $\widehat{\delta}_1$, multivariate linear regression analyses are applied. Results are reported in Table 3. The estimate[13] of the interaction term $Time \times Treated$ is significantly different from 0[14]. Accordingly, the development of mean OCB from the pre-TWSA to the post-TWSA period is significantly different between employees in regulated organizations compared to employees in unregulated organizations c.p., because the TWSA reduces OCB. Contrary to hypothesis 1, pay transparency caused by the TWSA unfolds a negative effect on OCB which remains stable across sets of control variables (Table 3).

Figure 2 provides an exploratory starting point for a first proposal on the potential moderation of the pay transparency-OCB relation by relative standing as posited in hypotheses 2 and 3. Figure 2 depicts a seemingly large difference in mean OCB of below and above median earners in the control group from the pre-TWSA period (before 2018) to the post-TWSA period (after 2018). Contrastingly, Figure 2 shows a relatively small difference in mean OCB of below and above median earners in the treatment group from before 2018 to after 2018. To derive statistically founded statements on hypotheses 2 and 3, algebraic analyses and multivariate linear regressions using OLS are applied. Formalized equations and regression results are shown and thematized in the following paragraph.

The model used to test hypotheses 2 and 3 extends the model introduced in (2) by adding a dummy variable for relative standing which indicates whether individuals earn less than a certain reference pay and interacting the newly introduced dummy with the existing variables

$$\begin{aligned}OCB_{itr} = \beta_0 + \beta_1 Relative\ standing_r \\ + \beta_2 Treated_i + \beta_3 Relative\ standing_r \\ \times Treated_i + \delta_0 Time_t + \delta_1 Time_t \\ \times Relative\ standing_r + \delta_2 Time_t \\ \times Treated_i + \delta_3 Time_t \times Relative\ standing_r \\ \times Treated_i + Controls_{itr} + u_{itr}. \end{aligned} \tag{3}$$

In (3), the triple interaction $\delta_3 Time \times Relative\ standing \times Treated$ constitutes the average treatment effect whereas the pairwise interactions serve as controls for the individual effects of time, relative standing, and organizational size. Algebraic rearrangements yield the DDD estimator δ_3

[12] The mathematical proof can be found in the appendix.

[13] Model 1.2 (Table 3): $\widehat{\delta}_{DD,Model1.2} = -0.184, p = 0.000147$

[14] Confidence level $\alpha = 0.001$

Table 1: Means and standard deviations

Variable	*M*	*SD*
1. OCB	3.85	0.69
2. Time	0.80	0.40
3. Treated	0.50	0.50
4. Relative standing calculated based on median pay	0.47	0.50
5. Relative standing calculated based on regressed reference pay	0.54	0.50
6. East/ West	0.77	0.42
7. Age	44.94	11.32
8. Sex	0.50	0.50
9. Wave	2.99	1.41
10. Education	3.74	0.93
11. Collective agreement	0.43	0.50
12. Size organization	231.42	139.98
13. Works council	0.46	0.50
14. HR department	0.91	0.29
15. Union member	0.16	0.37
16. Leadership role	0.36	0.48
17. Women percentage	43.09	17.70
18. Industry	8.56	3.96
19. Request	0.02	0.16
20. Knows the TWSA	0.25	0.43
21. Knows the term gender pay gap	0.27	0.44
n = 5,028		

Note: M and *SD* are used to represent mean and standard deviation, respectively.

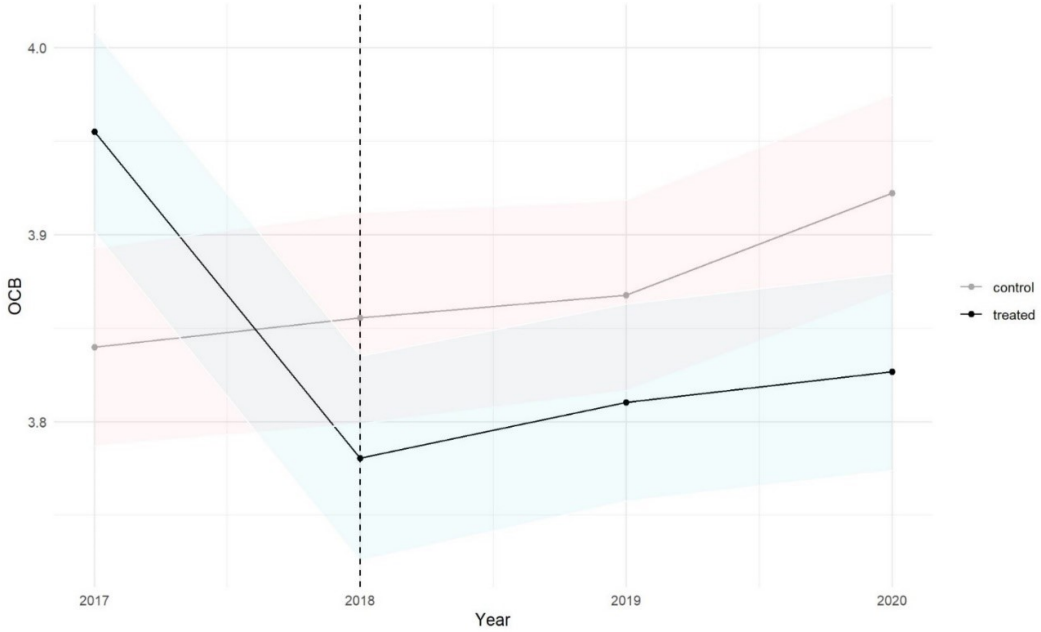


Figure 1: Visualization of Model 1.1 (Table 3) with 90% confidence intervals

Notes: Controls are omitted. Hence, the depicted values only match the estimates of Model 1.1 (Table 3) without controls.

Table 2: Correlations

Variable	1	2	3	4	5	6	7	8
1. OCB	(0.78)							
2. East/ West	-.00							
3. Age	.07***	.01						
4. Sex	.02	-.02	-.18***					
5. Wave	-.01	-.02	-.07***	-.02				
6. Education	.01	-.04***	-.13***	-.00	.05***			
7. Collective agreement	.09***	.04***	.07***	-.06***	.04***	-.00		
8. Size organization	-.02	.01	.05***	-.05***	.02	-.01	.18***	
9. Works council	-.00	.06***	.02	-.09***	.04**	.05***	.34***	.27***
10. HR department	.04***	.04***	-.07***	-.05***	.06***	.05***	.14***	.11***
11. Union member	.02	.02	-.00	-.08***	-.01	-.02	.24***	.11***
12. Leadership role	.14***	.00	.11***	-.18***	.01	.31***	.08***	.05***
13. Women percentage	.02	-.07***	-.01	.21***	.00	.06***	-.01	-.02
14. Industry	.00	-.07***	.01	.17***	-.02	.08***	-.04***	-.02
15. Request	-.01	.01	-.07***	-.04***	.05***	.04***	.11***	.15***
16. Knows the TWSA	.04***	.05***	-.03	-.09***	.06***	.18***	.08***	.07***
17. Knows the term gender pay gap	-.02	.05***	-.10**	-.03	.11**	.36**	.02	.02
Variable	9	10	11	12	13	14	15	16
10. HR department	.19***							
11. Union member	.26***	.07***						
12. Leadership role	.04***	.07***	.04***					
13. Women percentage	-.07***	-.02	-.08***	-.01				
14. Industry	-.09***	-.04***	-.08***	-.02	.75***			
15. Request	.11***	.03**	.16***	.07***	-.05***	-.03**		
16. Knows the TWSA	.08***	.06***	.09***	.16***	-.03**	-.01	.14***	
17. Knows the term gender pay gap	.02	.06***	-.01	.15***	-.00	.02	.07***	.31***
n = 5,028								

Notes: $^{**}p < 0.05$, $^{***}p < 0.01$. The value in parentheses denotes Cronbach's alpha.

$$\begin{aligned}\widehat{\delta}_3 = &\left[\left(\overline{OCB}_{\,Ti=1,\ Tr=1,RS=1} - \overline{OCB}_{\,Ti=0,\ Tr=1,RS=1}\right)\right.\\ &\left.-\left(\overline{OCB}_{\,Ti=1,\ Tr=0,RS=1} - \overline{OCB}_{Ti=0,\ Tr=0,RS=1}\right)\right]\\ &-\left[\left(\overline{OCB}_{Ti=1,Tr=1,RS=0} - \overline{OCB}_{Ti=0,Tr=1,RS=0}\right)\right.\\ &\left.-\left(\overline{OCB}_{Ti=1,Tr=0,RS=0} - \overline{OCB}_{Ti=0,Tr=0,RS=0}\right)\right]\\ &= \widehat{\delta}_{DD,RS=1} - \widehat{\delta}_{DD,RS=0} = \widehat{\delta}_{DDD}.\end{aligned} \tag{4}$$

Note: Time is abbreviated with Ti, Treated with Tr, and Relative standing with RS.

Similar to (1), the term in the first square bracket of (4) represents a DD estimator but only applied to observations that earn less than their reference pay. As control group, the first DD estimator uses individuals who are employed in organizations with 200 or less workers and earn less than their reference pay. Accordingly, $\widehat{\delta}_{DD,RS=1}$ represents the influence of the TWSA on OCB among employees who earn less than their reference pay. Opposingly, the term in the second square bracket of (4) contains $\widehat{\delta}_{DD,RS=0}$ which denotes the influence of the TWSA on OCB among workers who earn the same as or more than their reference pay. Obtaining the DD estimators and subsequently computing the difference yields the DDD estimator $\widehat{\delta}_{DDD}$ explicated in (4) which indicates the average difference in the effect of the TWSA on pay transparency due to relative standing. Further, OLS is applied to (3) to compute $\widehat{\delta}_{DDD}$ which corresponds to the coefficient of $Time \times Relative\ standing \times Treated$ reported in Table 4. Further, Table 4 includes two approaches to the determination of the dummy variable *Relative standing*. Whereas Models 2.1-2.3 assign values to *Relative standing* by using the median yearly gross pay of the sample, the second approach used for Models 3.1 – 3.3 defines an individual's peer group for social comparison via characteristic similarities (e.g. age, industry, position). Yet, the triple interaction of $Time \times Relative\ standing \times Treated$ Model 2.2 (Table 4): $\widehat{\delta}_{DDD,Model2.2} = 0.110, p_{DDD,Model2.2} = 0.252$; model 3.2 (Table 4): $\widehat{\delta}_{DDD,Model3.2} = 0.120, p_{DDD,Model3.2} = 0.223$ remains statistically insignificant in both approaches across all control variable specifications included in Table 4. The results propose that c.p. the difference in the average effect of

Table 3: Regression results of the DD with Time×Treated triple interaction

	Dependent variable:		
	OCB		
	Model 1.1	Model 1.2	Model 1.3
Constant	3.840***	3.587***	3.627***
	(0.000)	(0.000)	(0.000)
Time	0.034	0.064*	0.078**
	(0.302)	(0.091)	(0.046)
Treated	0.115***	0.078*	0.108**
	(0.008)	(0.069)	(0.049)
Time × Treated	-0.190***	-0.184***	-0.185***
	(0.000)	(0.000)	(0.000)
Wave = 2		-0.056*	-0.061*
		(0.072)	(0.051)
Wave = 3		-0.080***	-0.086***
		(0.009)	(0.006)
Wave = 4[a]		-0.047	-0.054*
		(0.118)	(0.076)
n	5,028	5,022	4,902
R^2	0.005	0.040	0.046
Adjusted R^2	0.004	0.035	0.039
Residual SE	0.692 (df = 5024)	0.681 (df = 4996)	0.680 (df = 4865)
F Statistic	8.091***	8.376***	6.546***
	(df = 3; 5024)	(df = 25; 4996)	(df = 36; 4865)
Controls	No	Partial[b]	Full[c]

Notes: *p < 0.1; **p < 0.05; ***p < 0.01. Values in parenthesis report p-values robust to heteroscedasticity.
[a]Wave 5 is used as reference category as time = 1 is identical to wave = 1.

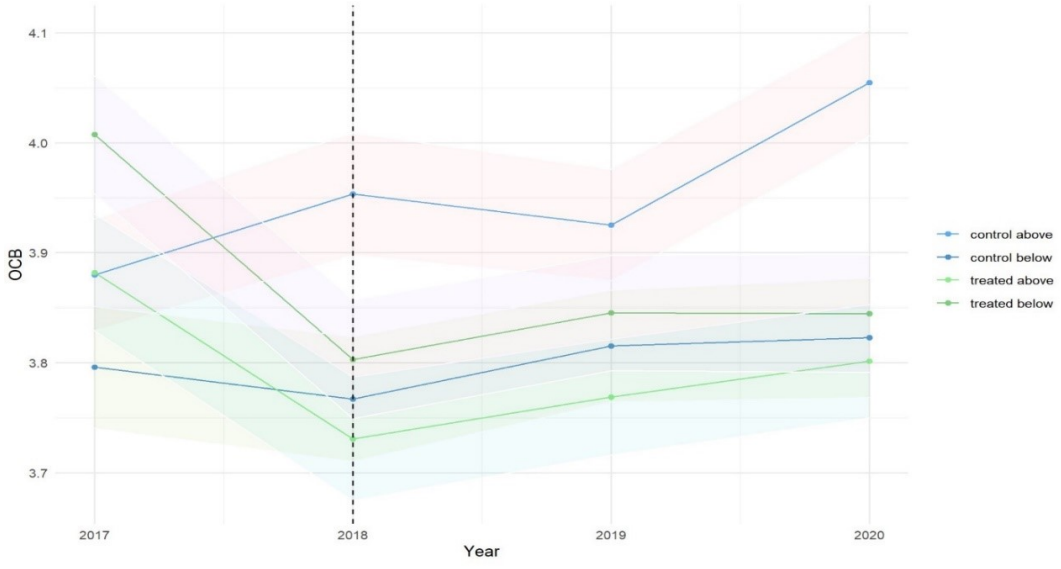


Figure 2: Visualization of Model 2.1 (Table 4) with 90% confidence intervals

Notes: Controls are omitted. Relative standing is determined via the median yearly gross pay of the sample. Hence, only the estimates of Model 2.1 (Table 4) match with the displayed values.

pay transparency on OCB is not significantly different from 0 between employees who earn less and employees who earn

Table 4: Regression results of the DDD with Time×Relative standing×Treated interaction

	Dependent variable:					
	OCB					
	Model 2.1	Model 2.2	Model 2.3	Model 3.1	Model 3.2	Model 3.3
Relative standing calculated using:	Median sample pay	Median sample pay	Median sample pay	Regressed reference pay	Regressed reference pay	Regressed reference pay
Constant	3.880***	3.637***	3.753***	3.809***	3.591***	3.657***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Treated	0.127**	0.096*	0.129*	0.185**	0.149**	0.183**
	(0.026)	(0.088)	(0.050)	(0.011)	(0.034)	(0.021)
Time	0.088*	0.107**	0.122**	0.098*	0.116**	0.146**
	(0.050)	(0.027)	(0.013)	(0.073)	(0.039)	(0.011)
Relative standing	-0.084	-0.029	-0.069	0.050	0.028	0.031
	(0.156)	(0.626)	(0.260)	(0.415)	(0.642)	(0.621)
Treated x Time	-0.254***	-0.239***	-0.237***	-0.261***	-0.256***	-0.267***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Time x Relative standing	-0.091	-0.073	-0.069	-0.111	-0.104	-0.136**
	(0.174)	(0.269)	(0.305)	(0.108)	(0.128)	(0.050)
Treated x Relative standing	-0.041	-0.041	-0.049	-0.115	-0.117	-0.129
	(0.636)	(0.639)	(0.572)	(0.203)	(0.188)	(0.152)
Treated x Time x Relative standing	0.118	0.110	0.105	0.116	0.120	0.136
	(0.226)	(0.252)	(0.286)	(0.249)	(0.223)	(0.173)
Wave = 2		-0.058**	-0.065**		-0.048	-0.051
		(0.039)	(0.040)		(0.122)	(0.102)
Wave = 3		-0.079***	-0.085***		-0.071**	-0.075**
		(0.006)	(0.006)		(0.021)	(0.018)
Wave = 4[a]		-0.047*	-0.054*		-0.041	-0.047
		(0.084)	(0.077)		(0.169)	(0.124)
n	5,028	5,022	4,902	5,028	5,022	4,902
R^2	0.014	0.042	0.051	0.007	0.043	0.050
Adjusted R^2	0.013	0.037	0.043	0.005	0.037	0.043
Residual SE	0.689	0.681	0.679	0.691	0.680	0.679
	(df = 5020)	(df = 4992)	(df = 4861)	(df = 5020)	(df = 4992)	(df = 4861)
F Statistic	10.308***	7.622***	6.474***	4.802***	7.687***	6.449***
	(df = 7; 5020)	(df = 29; 4992)	(df = 40; 4861)	(df = 7; 5020)	(df = 29; 4992)	(df = 40; 4861)
Controls	No	Partial[b]	Full[c]	No	Partial[b]	Full[c]

Notes: ∗p < 0.1; ∗p < 0.05; ∗∗∗p < 0.01. Values in parenthesis report p-values robust to heteroscedasticity.
[a]Wave 5 is used as reference category as time = 1 is identical to wave = 1.

as much as or more than their reference pay. Because relative standing does not significantly moderate the link between pay transparency and OCB, hypotheses 2 and 3 are rejected[15].

5. Discussion

5.1. Discoveries

Hypothesis 1 theoretically derives a positive effect of pay transparency on OCB based on social exchange theory. Yet, the empirical analyses find a significant negative effect. Drawing from research by Göbel et al. (N.d., p. 3) conducted with complementary plant-level data, organizations under the transparency condition did not actively increase pay information disclosure to provide employees with more pay transparency than required by the TWSA. Rather, pay information disclosure and the resulting pay transparency was kept at the legally required minimum or circumvented by organizations as the TWSA does not specify sanctions in case of violations (Göbel et al., n.d., p. 3; Weller & Göbel, 2019, pp. 21–22). In addition, only few individuals

[15]Confidence level $\alpha = 0.1$

enacted their right to request the median pay plus up to two wage components of an opposite-sex employee in a comparable position via the TWSA. The reason for little usage of the TWSA potentially stems from the constrained informativeness and incompleteness of the obtainable information which does neither allow inferences on the pay distribution of same-sex employees nor enable a comparison of the pay distribution between genders (Weller & Göbel, 2019, p. 22). Besides, employees are unlikely to completely know their rights due to the complexity of the TWSA requests (Weller & Göbel, 2019, p. 22). Possible circumventions by employers due to missing sanctions further hinder employees to exercise their right of information. However, public discussions and media coverage raised the expectation among employees that the TWSA would come along with increased pay transparency within organizations (Göbel et al., n.d., p. 22; Weller & Göbel, 2019, p. 23). Thus, employees may have implicitly added the expected increase in pay transparency to the psychological contract with their employers in case of a regulation by the TWSA. Yet, the actual implementation of the legislative change across affected organizations does not match with the aspiration raised by the government and reinforced by the media due to the TWSA's constrained informative value and resultingly little operational efficacy (Weller & Göbel, 2019, p. 22). Additionally, organizations do not voluntarily close the gap between the TWSA's aspired and actual effect by going beyond the legally required pay information disclosure (Göbel et al., n.d., p. 3). Rather, employees are left with uninformative and incomplete pay-related information or no information when organizations violate the TWSA (Weller & Göbel, 2019, pp. 22–23). Accordingly, the potentially positive effect of pay transparency on OCB as theorized in hypothesis 1 fails to materialize. The weak institutional pressure emitted by the TWSA combined with a lack of discretionary organizational support may lead to a disintegration of pay transparency as expected by employees and pay transparency as operationalized by organizations. The occurring gap between aspiration and reality may cause employees to perceive that the organization did not fulfill its part of the psychological contract. This theorizing is consistent with the findings by Göbel et al. (n.d., pp. 3–4) showing that employees of regulated plants increase the exchange of pay-related information among each other in search of an alternative source of information due to the non-informativeness of the TWSA combined with organizational passiveness to compensate this lack. Following Blau's (1984, p. 1) notion of social exchange, organizational failure to meet employees' pay transparency expectations via disclosure of pay-related information by exceeding the formally required minimum induces individuals to perceive their relationship with the organization as unbalanced on the social channel. Accordingly, employees reciprocate organizational behavior by reducing their discretionary OCB to rebalance the social exchange with the organization in a tit-for-tat manner.

Concerning hypotheses 2 and 3, no significant moderating effect of relative standing on the relation between pay transparency and OCB derived from social comparison theory (Festinger, 1954, p. 117) and attribution theory (Miller & Ross, 1975, p. 213) was found. A possible explanation for the insignificance may be the restricted informativeness of requests based on the TWSA as noted earlier. The TWSA may not provide employees with sufficient information to cause social comparison which would allow a (more precise) determination of relative standing on the pay distribution among peers so that the resulting psychological attribution of the outcome would significantly influence the link between pay transparency and OCB. Accordingly, the prevalent critique that pay transparency leads to detrimental behavioral outcomes due to social comparison processes (Cullen & Perez-Truglia, 2018, pp. 4–6; Dube et al., 2019, p. 639) is not fundamentally acceptable. The statistical insignificance of relative standing as moderator implies that the pay transparency-OCB relation does not significantly change with regard to relative standing in case of selective pay transparency as created by the TWSA.

5.2. Academic and managerial implications

This study investigates the effect of selective pay transparency as created by the TWSA on OCB. Related studies tend to focus on the extremes of the continuum spanned by pay transparency and pay secrecy (Brown et al., 2022, p. 10). This paper constitutes a step towards a more nuanced approach to pay transparency. By reporting a negative effect of pay transparency on OCB contrary to the proposal by Marasi et al. (2018, p. 58), the investigation strengthens the suggestion by SimanTov-Nachlieli and Bamberger (2021, p. 230) that varying forms of transparency may lead to differential outcomes. Hence, this study contributes to a progressive coherence of pay transparency research. Also, the common critique that pay transparency negatively impacts employees' behavior because of social comparison is not supported in case of selective pay transparency as investigated in this study. Besides, this research integrates social exchange theory (Blau, 1984, p. 1), social comparison theory (Festinger, 1954, p. 117), and attribution theory (Miller & Ross, 1975, p. 213) to obtain a more sophisticated understanding of the socio-psychological mechanisms which are induced by pay transparency and influence behavioral outcomes displayed by individuals.

The paper also provides managerial implications. Usually, employers possess more pay-related knowledge than their employees. This creates one-sided perfect information in favor of organizations (Stiglitz, 1985, p. 24). If the underinformed actor is interested in the disclosure of pay information, the inferior party is likely to seek for a reduction of pay-related information asymmetry through interacting with the organization as superiorly informed sender and thus achieve a pay transparency equilibrium (Brown et al., 2022, p. 5). Due to a stronger relative position within the sender-receiver interaction, organizations can design pay communication policies and practices to steer the employees' knowledge of pay-related information. Consequently, a

company's pay information disclosure strategy can either facilitate or hamper pay transparency and thus influence OCB of employees. The results of this study imply that disclosing a minimum of pay-related information characterized by little informativeness or circumventing disclosure may not be an optimal response to a regulatory change towards more pay transparency. Regarding the introduction of the TWSA in Germany, such strategies result in a detrimental effect of pay transparency on OCB as the operationalized pay transparency may not be consistent with employees' expectations raised by institutional forces and media coverage.

5.3. Limitations and future research proposals

The prevalent study remains limited by the data used for the empirical analysis. Using the sample median of yearly gross pay fails to define a peer group for social comparison. The human capital variables used to estimate a reference pay via regression only approximately define an individual's peer group. Accordingly, the results used to test hypotheses 2 and 3 may be diluted. To define a peer group more precisely, team level data would be needed. Also, choosing an arbitrary cut off for the dichotomization of the relative standing dummy leads to an assignment of the value 0 to observations with a salary equal to reference pay. This might marginally blur effects associated with relative standing. Moreover, the causal interpretation conducted in section 4 assumes parallel trends in the control and treatment group as well as stable unit treatment values. Yet, the parallel trends assumption underlying the DD and DDD cannot be tested because only one wave of data was collected before the introduction of the TWSA. Also, this study interprets results of the DDD causally, despite between-group variance. Further, the findings regarding pay transparency are based on the introduction of the TWSA as transparency condition. Accordingly, this study investigates pay transparency as created by the TWSA. Hence, transferability of the findings to other settings may be limited because different forms of pay transparency created via laws with specifications different from those of the TWSA for instance may lead to differential outcomes than reported in this study.

Besides, this study is limited by its scope. The mechanisms used to connect pay transparency and OCB remain on a theoretical level and thus require empirical testing to develop a more sophisticated understanding of how the pay transparency-OCB relation works. Further research is also needed to examine how various dimensions of pay information disclosure and the resulting levels of pay transparency differ in their influence on employees' attitudes and behaviors. Therefore, structural equation modelling could be used to examine the weights on the links between the dimensions of pay information disclosure, resulting forms of pay transparency, attitudinal (e.g. motivation, commitment), and behavioral outcomes (e.g. turnover, OCB). Building on research by Göbel et al. (n.d., p. 22), the TWSA creates selective top-down pay transparency but also positive externalities on emergent pay transparency. This research does not further distinguish the influence of pay transparency on OCB by channels used to obtain pay-related information. Accordingly, examining whether top-down transparency and bottom-up transparency created through the exchange of pay-related information among employees induce differing behavioral consequences may provide fruitful ground for future investigations. Hereof, disaggregating the latent concept of OCB into OCB directed at individuals and OCB addressed at the organization may deliver an answer to the question whether the addressee of the change in OCB corresponds with the sender of pay-related information. Lastly, coding relative standing as metric instead of dichotomous variable as conducted by SimanTov-Nachlieli and Bamberger (2021, p. 237) prevents loss of information and would allow future studies to investigate how attitudes and behaviors may change over an increasing distance to the reference pay.

6. Conclusion

This paper investigates the understudied relation between pay transparency and OCB as well as a potential moderation by relative standing. Building on social exchange, social comparison, and attribution theory, theoretical arguments for a positive effect of pay transparency on OCB and a moderating role of relative stranding are developed. However, using a policy change (TWSA) as transparency condition in a quasi-experiment, the empirical analyses find a significant negative effect of pay transparency on OCB, whereas the moderation by relative standing is rejected. Resultingly, this research adds to the prevalent pay transparency literature by clarifying the effect of selective pay transparency as created by the TWSA on OCB which differs from theoretical predictions and thus highlights the need for a more nuanced approach to different forms of pay transparency. Further, a common critique of pay transparency is addressed by testing relative standing as moderator with insignificant results. To mitigate the negative effect of pay transparency on OCB, this study proposes that managers refrain from pay information disclosure strategies which provide employees with no or little meaningful pay-related information as reaction to legislative changes towards more pay transparency such as the TWSA in Germany.

References

Austin, W., McGinn, N. C., & Susmilch, C. (1980). Internal standards revisited: Effects of social comparisons and expectancies on judgments of fairness and satisfaction. *Journal of Experimental Social Psychology*, *16*(5), 426–441.

Bamberger, P., & Belogolovsky, E. (2017). The dark side of transparency: How and when pay administration practices affect employee helping. *Journal of Applied Psychology*, *102*(4), 658–671.

Belogolovsky, E., & Bamberger, P. A. (2014). Signaling in secret: Pay for performance and the incentive and sorting effects of pay secrecy. *Academy of Management Journal*, *57*(6), 1706–1733.

Blau, P. M. (1984). *Exchange and power in social life*. New York: J. Wiley and Sons.

Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality∗. *The Quarterly Journal of Economics*, *133*(2), 611–663.

Brown, D. J., Ferris, D. L., Heller, D., & Keeping, L. M. (2007). Antecedents and consequences of the frequency of upward and downward social comparisons at work. *Organizational Behavior and Human Decision Processes*, *102*(1), 59–75.

Brown, M., Nyberg, A. J., Weller, I., & Strizver, S. D. (2022). Pay information disclosure: Review and recommendations for research spanning the pay secrecy–pay transparency continuum. *Journal of Management*, *48*(6), 1661–1694.

Bundesministerium für Familie, Senioren, Frauen und Jugend. (2017). *Das Entgelttransparenzgesetz: Informationen zum Gesetz zur Förderung der Entgelttransparenz.* https://www.bmfsfj.de/resource/blob/117322/c9ef7c4bbe4822e644c94821b09aa88f/das-entgelttransparenzgesetz-informationen-zum-gesetz-zur-foerderung-der-entgelttransparenz-data.pdf.

Bundesministerium für Justiz. (2022). *EntgTranspG: Gesetz zur Förderung der Entgelttransparenz zwischen Frauen und Männern.* https://www.gesetze-im-internet.de/entgtranspg/BJNR215210017.html.

Burns, T., & Stalker, G. M. (2001). *The management of innovation*. Oxford University PressOxford.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, *102*(6), 2981–3003.

Castilla, E. J. (2015). Accounting for the gap: A firm study manipulating organizational accountability and transparency in pay decisions. *Organization Science*, *26*(2), 311–333.

Colella, A., Paetzold, R. L., Zardkoohi, A., & Wesson, M. J. (2007). Exposing pay secrecy. *Academy of Management Review*, *32*(1), 55–71.

Colquitt, J. A., Conlon, D. E., Wesson, M. J., Porter, C. O. L. H., & Ng, K. Y. (2001). Justice at the millennium: A meta-analytic review of 25 years of organizational justice research. *Journal of Applied Psychology*, *86*(3), 425–445.

Colquitt, J. A., LePine, J. A., Piccolo, R. F., Zapata, C. P., & Rich, B. L. (2012). Explaining the justice–performance relationship: Trust as exchange deepener or trust as uncertainty reducer? *Journal of Applied Psychology*, *97*(1), 1–15.

Cullen, Z., & Perez-Truglia, R. (2018). *How much does your boss make? the effects of salary comparisons* (Tech. Rep.).

Day, N. E. (2011). Perceived pay communication, justice and pay satisfaction. *Employee Relations*, *33*(5), 476–497.

Dreitzel, H. P. (1980). *Die gesellschaftlichen Leiden und das Leiden an der Gesellschaft: Eine Pathologie des Alltagslebens* (3rd ed.). Stuttgart: Ferdinand Enke Verlag.

Dube, A., Giuliano, L., & Leonard, J. (2019). Fairness and frictions: The impact of unequal raises on quit behavior. *American Economic Review*, *109*(2), 620–663.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.

Frank, R. H. (1985). The Demand for Unobservable and Other Nonpositional Goods. *The American Economic Review*, *75*(1), 101–116.

Fulmer, I. S., & Chen, Y. (2014). How communication affects employee knowledge of and reactions to compensation systems. In V. D. Miller & M. E. Gordon (Eds.), (chap. Meeting the Challenge of Human Resource Management: A Communication Perspective). New York: Routledge.

Futrell, C. M., & Jenkins, O. C. (1978). Pay secrecy versus pay disclosure for salesmen: A longitudinal study. *Journal of Marketing Research*, *15*(2), 214.

Gupta, N., & Shaw, J. D. (2014). Employee compensation: The neglected area of HRM research. *Human Resource Management Review*, *24*(1), 1–4.

Göbel, L., Weller, I., & Nyberg, A. J. (n.d.). *How employers and employees react to rising pay transparency expectations: An exploratory study*. Manuscript submitted for publication.

Göbel, L., Weller, I., & Nyberg, A. J. (2020). How employers and employees react to rising pay transparency expectations: An exploratory study. *Academy of Management Proceedings*, *2020*(1), 17109.

Heider, F. (1958). *The psychology of interpersonal relations.* John Wiley & Sons Inc.

Johnson, D. D. P., & Fowler, J. H. (2011). The evolution of overconfidence. *Nature*, *477*(7364), 317–320.

Katz, D. (1964). The motivational basis of organizational behavior. *Behavioral Science*, *9*(2), 131–146.

Kelley, H. H. (1971). *Attribution in social interaction*. New York: General Learning Press.

Konovsky, M. A., & Organ, D. W. (1996). Dispositional and contextual determinants of organizational citizenship behavior. *Journal of Organizational Behavior*, *17*(3), 253–266.

Konovsky, M. A., & Pugh, S. D. (1994). CITIZENSHIP BEHAVIOR AND SOCIAL EXCHANGE. *Academy of Management Journal*, *37*(3), 656–669.

Langer, E. J., & Roth, J. (1975). Heads i win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology*, *32*(6), 951–955.

Lawler, E. E. (1965). Managers' perceptions of their subordinates pay and of their superiors' pay. *Personnel Psychology*, 413–422.

Lawler, E. E. (1967). Secrecy about management compensation: Are there hidden costs? *Organizational Behavior and Human Performance*, *2*(2), 182–189.

Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *Journal of Applied Psychology*, *87*(1), 131–142.

LePine, J. A., Erez, A., & Johnson, D. E. (2002). The nature and dimensionality of organizational citizenship behavior: A critical review and meta-analysis. *Journal of Applied Psychology*, *87*(1), 52–65.

MacKenzie, S. B., Podsakoff, P. M., & Fetter, R. (1993). The impact of organizational citizenship behavior on evaluations of salesperson performance. *Journal of Marketing*, *57*(1), 70–80.

Marasi, S., & Bennett, R. J. (2016). Pay communication: Where do we go from here? *Human Resource Management Review*, *26*(1), 50–58.

Marasi, S., Wall, A., & Bennett, R. J. (2018). Pay openness movement: Is it merited? does it influence more desirable employee outcomes than pay secrecy? *Organization Management Journal*, *15*(2), 58–77.

Matiaske, W., Wallmeier, G., & Weller, I. (2017). Rollen, Extra-Rollenverhalten und Organizational Citizenship Behavior. In A. Martin (Ed.), (2nd ed., pp. 254–279). Stuttgart: Kohlhammer Verlag.

Matiaske, W., & Weller, I. (2007). Do extrinsic rewards enhance organizational citizenship behavior? a study of public sector organizations. In *Public governance and leadership* (pp. 513–534). Wiesbaden: DUV.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213–225.

Montag-Smit, T. A., & Smit, B. W. (2021). What are you hiding? employee attributions for pay secrecy policies. *Human Resource Management Journal*, *31*(3), 704–728.

Nienhüser, W. (1993). Rolle. In W. Weber, W. Mayrhofer, & W. Nienhüser (Eds.), (chap. Sammlung Poeschel: Bd. 127. Grundbegriffe der Personalwirtschaft). Stuttgart: Schäffer-Poeschel.

Organ, D., Podsakoff, P., & MacKenzie, S. (2006). *Organizational citizenship behavior: Its nature, antecedents, and consequences*. SAGE Publications, Inc.

Organ, D. W. (1988). *Organizational citizenship behavior: The good soldier syndrome. issues in organization and management series*. Lexington, MA: D.C. Heath and Company.

Organ, D. W., & Ryan, K. (1995). A META-ANALYTIC REVIEW OF ATTITUDINAL AND DISPOSITIONAL PREDICTORS OF ORGANIZATIONAL CITIZENSHIP BEHAVIOR. *Personnel Psychology*, *48*(4), 775–802.

Podsakoff, P. M., MacKenzie, S. B., Paine, J. B., & Bachrach, D. G. (2000). Organizational citizenship behaviors: A critical review of the theoretical and empirical literature and suggestions for future research. *Journal of Management*, *26*(3), 513–563.

Rotundo, M., & Sackett, P. R. (2002). The relative importance of task, citizenship, and counterproductive performance to global ratings of job performance: A policy-capturing approach. *Journal of Applied Psychology*, *87*(1), 66–80.

Rousseau, D. (1995). *Psychological contracts in organizations: Understanding written and unwritten agreements*. SAGE Publications, Inc.

Schmidt, J. (2017). *Unerwünschte Effekte von Lohntransparenz?* (Tech. Rep.). IW-Report. (21), 1–41. Retrieved from https://www.iwkoeln.de/studien/joerg-schmidt-unerwuenschte-effekte-von-lohntransparenz.html

Schnackenberg, A. K., & Tomlinson, E. C. (2016). Organizational transparency. *Journal of Management*, *42*(7), 1784–1810.

Schnake, M., & Dumler, M. P. (1997). Organizational citizenship behavior: The impact of rewards and reward practices. *Journal of Managerial Issues*, *9*, 216–229.

SimanTov-Nachlieli, I., & Bamberger, P. (2021). Pay communication, justice, and affect: The asymmetric effects of process and outcome pay transparency on counterproductive workplace behavior. *Journal of Applied Psychology*, *106*(2), 230–249.

Smit, B. W., & Montag-Smit, T. (2019). The pay transparency dilemma: Development and validation of the pay information exchange preferences scale. *Journal of Applied Psychology*, *104*(4), 537–558.

Smith, C. A., Organ, D. W., & Near, J. P. (1983). Organizational citizenship behavior: Its nature and antecedents. *Journal of Applied Psychology*, *68*(4), 653–663.

Stiglitz, J. E. (1985). Information and economic analysis: A perspective. *The Economic Journal*, *95*, 21.

van den Bos, K., & Lind, E. A. (2002). Uncertainty management by means of fairness judgments. In *Advances in experimental social psychology* (pp. 1–60). Elsevier.

Veldman, A. (2017). *Pay transparency in the EU: a legal analysis of the situation in the EU Member States, Iceland, Liechtenstein and Norway.* European Commission Publications Office.

Wagner, S. L., & Rush, M. C. (2000). Altruistic organizational citizenship behavior: Context, disposition, and age. *The Journal of Social Psychology*, *140*(3), 379–391.

Weller, I., & Göbel, L. (2019). Ein Jahr Entgelttransparenzgesetz. Das Gegenteil von gut ist gut gemeint. *Ifo Schnelldienst*(72), 21–26.

Weller, I., Matiaske, W., & Holtmann, D. (2007). Leistungsorientierte entlohnung, extra-rollenverhalten und commitment. In M. Moldaschl (Ed.), (2nd ed., chap. Arbeit, Innovation und Nachhaltigkeit: Vol. 3. Immaterielle Ressourcen: Nachhaltigkeit von Unternehmensführung und Arbeit I). München: Hampp.

Wooldridge, J. M. (2019). *Introductory econometrics* (7th ed.). Mason, OH: Cengage.

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality*, *47*(2), 245–287.

Junior Management Science 8(3) (2023) 732-771

Junior Management Science

journal homepage: www.jums.academy

A Techno-Economic Analysis of Space-Based Solar Power Systems

Benedikt Kruft

Technische Universität München

Abstract

Space-based solar power (SBSP) promises to provide flexible renewable baseload power. However, no full-system prototype exists due to a perceived lack of economic viability. The goal of this thesis is therefore to determine how different technology approaches can improve key technical metrics of SBSP and consequently the economics. For this purpose, we divide the system into its three main segments and define critical metrics for the performance of each subsystem. Based on these, novel technology approaches from the literature are then evaluated. For the solar satellite, we are able to show that a number of technology options exist that might improve power levels, radiation resistance, and mass-related ratios. These advances would greatly benefit overall system economics, as the space segment constitutes a big lever for enhancing the levelised cost of electricity (LCOE). Furthermore, microwave power beaming efficiencies in line with required levels have been demonstrated but so far lack the scale and distance necessary for SBSP. Ultimately, the global capacity in space lift capabilities appears to be a major bottleneck. Consequently, a reduction in mass of the satellite would not only be a matter of economics but might render any such project even possible in the first place.

Keywords: Energy; Solar; Space; Microwaves; Sustainability.

1. Introduction

The goal of this thesis is the techno-economic analysis of a general space-based solar power (SBSP) system and its subsystems. We do not focus on one of the many proposed designs in particular, but rather use a selection of the most advanced ones as a point of reference to potentially identify universal levers for progress. So far, the literature has mostly focussed on determining the general theoretical feasibility of SBSP and that of specific concepts in their entirety, such as the CASSIOPeiA project. As a result, we observed a lack in granularity that would enable the evaluation of segment-specific technology approaches. The reason for this might be the prevailing perception of SBSP as a niche and moonshot energy solution. Therefore, a targeted analysis that aims to compare technological solutions to problems specific to SBSP based on the critical metrics of each subsegment is so far missing. Consequently, our research question is how different technology approaches can improve technical metrics of SBSP subsystems and in turn help overall system economics.

To achieve this, we start by briefly describing the context of the global energy transition and which role SBSP could play. Here we illustrate the basic idea behind SBSP plus some of the key benefits the technology has to offer. However, a comparison with other renewable generation capacities is outside the scope of this thesis. Next, we describe the concept of SBSP in more detail, including some of the most relevant modern designs under development. We then establish a general model of a SBSP system and divide it into three subsystems for further differentiation.

The division into space segment, wireless power transmission (WPT) and ground structure, as well as space launch and infrastructure then serves as the structure for our techno-economic analysis. For each of these segments, we start by identifying the metrics most critical to the technical and economic performance of the respective subsystem. This is followed by an evaluation of different technology options and developments for all subsystems by these metrics. Particularly for the space segment, each technological alternative

I want to thank the Entrepreneurial Masterclass program by UnternehmerTUM that gave me the opportunity to write about this topic and all participants who provided me with sparring opportunities and guidance. Furthermore, I want to thank Prof. Dr. Thomas Hamacher for supervising my thesis. Lastly, I am gratefull for the proof-reading support provided by my family and friends.

DOI: https://doi.org/10.5282/jums/v8i3pp732-771

poses a potential solution to one or multiple problems encountered during SBSP development. The selection of analysed alternatives should not be considered as exhaustive, but rather focusses on innovations that appear to be the most promising and impactful for SBSP. When discussing WPT, we will also emphasise some of the historical developments and demonstrations that have led to where the technology stands today. For the third and last segment, we will focus on the launch infrastructure in particular and only briefly discuss some production-related aspects. The three segments are concluded by a respective overview of the established alternatives or developments and an outlook on which technology combinations appear to be the most promising.

To establish equality for the comparisons of metrics drawn from the literature, some had to be recalculated or adjusted based on the disclosed data. To ensure metrics are adjusted following shared rules, we establish a common set of formulas as well as adjustment factors which are applied equally. Overall, the resulting metrics for different subsystem configurations should be seen as a heuristic approach to achieve the aforementioned comparability of different technology options. Therefore, they should be interpreted as an indicator for the order of magnitude of the impact of these technology options. All calculations are performed in a separate model, which is available upon request.

Throughout this thesis, we will also sometimes refer to some of the analyses from the literature on levelised cost of electricty (LCOE). Given how incomparable many of the absolute numbers are due to being based on sometimes vastly differing assumptions, we mostly use any breakdowns regarding cost contributions by the subsystems as a relative point of reference. This allows us to more accurately estimate how big the economic impact of the previously discussed technological levers could be, relative to each other. Consequently, this thesis does not conduct in-depth cost modelling or calculations of its own. Policy, geopolitical, and social considerations, which could impact LCOE, are also outside the scope of this thesis. It is furthermore not the purpose of this thesis to compare LCOE numbers between different renewable technologies and SBSP.

We will also introduce sustainability-related metrics such as emissions per kWh and energy payback times in the last chapter. However, a full environmental analysis of SBSP is outside the scope of this thesis. Instead, our goal is to gain a first understanding of the direction in which current approaches are trending in terms of environmental impacts and energy use. These insights can then serve as the basis for further research.

Finally, we sum up our findings, including projected LCOE numbers on a system level, and their consequences for the development of commercial SBSP systems going forward.

1.1. Space-based solar power for the global energy transition

The most recent IPCC report has once again made clear that quick and far-reaching action is needed if we are to have a chance at limiting global warming to 1.5 degrees (IPCC, 2022). To achieve global net zero by 2050, widespread electrification is a crucial step as it allows us to decarbonise many industries at once by turning our electricity generation green.

So far, a selection of renewable energy sources has been accessed through large amounts of investment. Solar energy has even become the cheapest energy source in history (IEA, 2020). Nonetheless, more financing is needed to overcome some of the remaining issues of our primary green energy sources, terrestrial solar and wind. Both of them suffer from intermittency, be it due to weather or day-night cycles. Furthermore, their generation is inflexible and difficult to predict far in advance. These are some of the reasons why they are generally labelled as non-dispatchable.

Space-based solar power (SBSP) promises to overcome this issue by becoming the first dispatchable renewable baseload power. The general idea is to place a satellite with extensive solar power capacity in orbit around earth. The sometimes with the help of additional mirrors and reflectors collected solar radiation is then turned into DC power before being converted to either a laser or radio frequency (RF) beam. This beam is then targeted at a receiving station on the ground, where the wirelessly transmitted power is reconverted first to DC then AC before being fed into the grid. A simplified representation of this process including efficiencies is provided in figure 1.

SBSP has a number of benefits, which are described at length in various publications (e.g. Jaffe, 2020). First, if a geostationary orbit (GEO) is chosen and the right array structure employed, the satellite would enjoy virtually constant illumination, except for during the spring and autumn equinox with a 70 min blackout each (Way & Lamyman, 2020). Hence, intermittency would no longer be an issue and the electricity generated by SBSP would be highly predictable far in advance. Second, the solar irradiation collected by the satellite would be unattenuated by weather or the atmosphere, giving a greater yield per unit surface area. Third, the beam would have to be steerable to enable precise pointing, which in turn allows flexibility in regards to where the energy is directed to. Multiple ground stations could be employed across different countries and even continents, which would then be fed by the same satellite or constellation of satellites in an alternating or semi-continuous fashion. Therefore, SBSP can also be used as a diplomatic or development tool. With mobile receiver stations, the beam could even be utilized to provide electricity to disaster-struck areas. Lastly, the technology promises to be sustainable over its lifetime, although more detailed studies will be necessary to confirm the precise extent. Overall, SBSP thus has the potential of providing semi-globally dispatchable renewable baseload power.

On the other hand, there are also some concerns associated with SBSP. Space is a domain that is difficult to control and therefore placing an asset critical to a nation's energy autonomy so far away from it comes with security concerns. Recent acts by Russia have added to this worry by showing that targeted destruction of satellites is well within the means of space-faring nations (The Economist, 2021b).

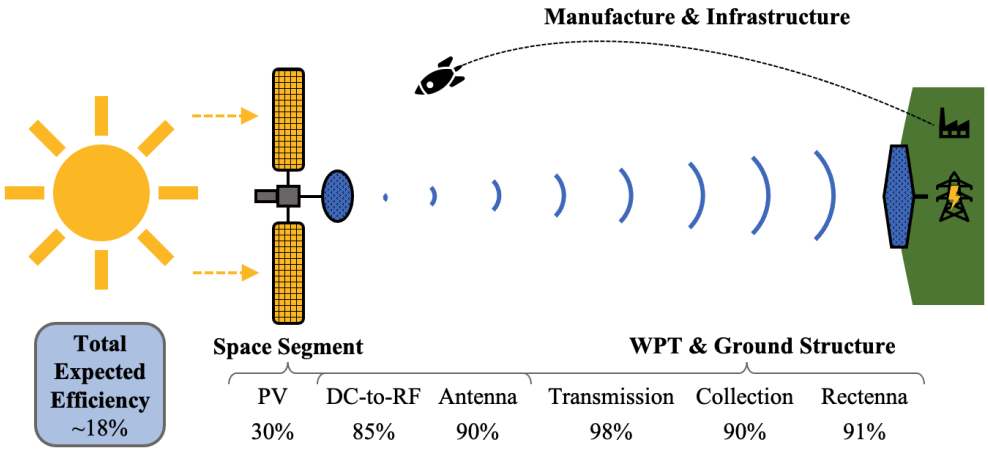


Figure 1: A general SBSP system generates solar power in space via a satellite which is then transformed and transmitted wirelessly, here in the form of microwaves, to a receiving station on the ground. The space segment would be manufactured on earth to be launched to and assembled in GTO. Overall, system efficiencies of 18% are expected, based on subsystem demonstrations. Own representation with efficiencies based on Jaffe (2020).

Furthermore, while reaching our climate targets requires fast and decisive action, SBSP is still not close to having any full-system prototype in orbit, thus putting in question whether it can be deployed fast enough to contribute meaningfully. Potentially, the considerable amount of funds required should then better be spent elsewhere. While the overall economics of solar power from space do appear to be competitive based on modelling (e.g. Madonna, 2018), those too would have be to confirmed in practice. Additionally, as we continue to launch ever more satellites, critical orbits such as GEO are starting to become congested (Euroconsult, 2021). Lastly, public acceptance and safety concerns around what are essentially kilometre-wide invisible energy beams will have to be addressed very early in the process.

The idea of SBSP was first described in Isaac Asimov's short story 'Reason' in 1941 and later introduced to academia by Peter Glaser (Glaser, 1968), who also filed the first SBSP patent (Glaser, 1973). What followed were a number of studies (e.g. Koomanoff, 1981) and critiques (e.g. Corson et al., 1981) of the concept. Particularly NASA and other U.S. government agencies investigated the idea in more depth, developing the two suntower concepts at 2.45 GHz (NASA, 1978) and 5.8 GHz (Davis, 2012), which are still sometimes used as reference systems. The private sector is also involved with a number of patents and plans for commercialization (Solaren, 2022).

1.2. Modern SBSP concepts

Today, there are a number of nations and organisations active in SBSP research and development. While for some actors, such as China, the primary goal is to develop a fully functioning concept, others focus more on advancing key technologies like WPT. For instance, Japan and the California Institute of Technology rather fall in the second category. SBSP satellites remain an ambitious idea, given that the most advanced modern space segment designs are measured in kilometres and weigh thousands of tons. In comparison, as the largest man-made structure in space and the result of a collaboration involving space agencies from 15 countries, the ISS is barely 100 m long and weighs just over 400 t (Garcia, 2021).

Today, a variety of concepts exist that have entered advanced planning stages. The ones we have chosen as reference systems during our comparative analyses were selected based on how established they are in the literature and the quantity as well as quality of data available to perform the necessary calculations. Furthermore, we have selected space segment architectures that are different from each other to capture the impact of design variations on the critical metrics. Lastly, all of the chosen concepts have some form of government backing, which will likely be essential given the scale and ambition of SBSP. Therefore, we want to highlight the following projects, whose respective unique space segment architecture is also displayed in figure 2.

SPS-ALPHA:
The Solar Power Satellite with Arbitrarily Large PHased Array (SPS-ALPHA) is a concept that has been developed by John Mankins since the 1990s in collaboration with NASA (Mankins, 1997). The about 4-km tall structure is comprised of an energy conversion array, which is connected to a single large reflector array via boom structures (Mankins, 2021). Once placed in GEO, it would maintain an orientation where the antenna side of the conversion array is continuously pointed at the receiving station on earth, transmitting the energy in form of microwaves at 2.45 GHz. During local night, the reflector array will redirect and concentrate the sunlight onto the photovoltaic (PV) surface to ensure continuous operation. At 7,600 t in-space mass, it is one of the heavier concepts. It is designed to deliver 2 GW of DC power to the grid

around the clock.

CASSIOPeiA:
The Constant Aperture, Solid-State, Integrated, Orbital Phased Array (CASSIOPeiA) concept has been developed by Ian Cash from International Electric. Its key feature is a 1.7-km long helical energy generation and transmission array with two conical solid-state symmetrical concentrators on either end (Cash, 2019). These primary reflectors collect, concentrate, and collimate the sunlight onto the patented helix array, which includes further concentration on the PV chips. L-shaped sandwich modules are employed to form the array with a 360-degree beam steering capability along the orbital plane. The satellite has no moving parts, weighs 2,000 t, and would deliver 2 GW of DC power to the grid from GEO via a 2.45 GHz microwave beam. CASSIOPeiA has also been investigated by the UK government as a way to de-risk their transition towards green electricity generation and looks likely to be pursued further after positive results from recent studies (Way & Lamyman, 2021b).

MR-SPS:
The Multi-Rotary joints Solar Power Satellite (MR-SPS) was first proposed in 2014 by China and recently updated in 2021 (Hou & Li, 2021). Out of our selected concepts, it is the only one that employs separated generation and transmission surfaces without any solar concentration. Therefore, it is comprised of freely rotating solar arrays to adjust to the sun's position. These are then connected via trusses with 100 rotary joints and cabling to a transmission antenna, which emits microwaves at 5.8 GHz from GEO. It is by far the heaviest concept at 10,000 t of in-space mass and generates 1 GW of DC power that can be fed into the grid.

Other notable concepts include older architectures by NASA and a number of designs by Japan. However, Japanese agencies have mostly transitioned to researching key technologies such as WPT and not published any updated concepts with parameters recently. The California Institute of Technology is moving in the opposite direction. Having initially succeeded in advancing high-performance ultralight sandwich modules for planar arrays, they are now starting to develop their own full-system concept Madonna (2021).

1.3. A general SBSP system model

To analyse different SBSP architectures from a technological and ultimately also an economic perspective, a common system model is required along which we can structure our approach. These defined subsystems also need to reflect the biggest levers for technological and ultimately cost improvements. Therefore, LCOE calculations from the literature based on subsegments (e.g. Marshall, Madonna, & Pellegrino, 2021) informed our decision on where to draw system boundaries.

Consequently, we will divide SBSP into a space segment, WPT and necessary ground structures, and manufacturing as well as infrastructure. These segments are also indicated in figure 1. The space segment contains all structures that are permanently placed in orbit, such as PV arrays, reflectors, propulsion systems, and antennas. The WPT subsystem also concerns the antenna, the transmission process, and the reception as well as reconversion of the beam. Lastly, we look at anything related to the construction, assembly, maintenance, and launch of the satellite.

Gauging their respective importance for and impact on LCOE, launch costs are repeatedly identified in the literature as one of the largest drivers in variance for system cost (e.g. Way & Lamyman, 2021b). Consequently, developments in the launch market will therefore receive particular focus in section 4. Additionally, total launch costs are also influenced by the maintenance requirements and satellite mass. Hence, we will focus some of our metrics for the space segment on mass- and lifespan-related performance. Furthermore, the mass of the space segment can also act as a proxy for manufacturing costs and therefore has a dual impact on overall costs. With these circumstances in mind, we can begin our analysis with the space segment.

2. Space segment

First, we focus on the satellite or space segment of SBSP systems. While the antenna transmission array is part of our definition of the satellite, it overlaps with the wireless power transfer subsystem and is discussed in more detail in section 3. Hence, the consideration of components pertaining to the transfer of wireless power, such as the transfer antenna apertures, will only extend to their interaction with the overall satellite structure and design. This includes to some extent DC-to-RF conversion efficiencies to calculate metrics based on the satellite's RF power output as well as the implications of building an array of sandwich panels, which are discussed in section 2.5. DC-to-RF efficiencies are then addressed more directly in the subsequent chapter. Due to the lack of laser based SBSP transmission systems in the literature, a dynamic further investigated during the chapter on power beaming, our analysis for the space segment will exclusively focus on microwaves as a transmission modality. Overall, particular focus is placed on the PV cells and concentrators, module design, satellite structure, and the overall array management. We will also briefly discuss the impact of different choices of orbit.

2.1. Metrics for space segment evaluation

When comparing metrics for the space segment, it is important to be clear on which measurements are being used. For example, some PV prototypes on which metrics are reported do not yet include the transmit antenna array, hence reducing overall mass. It is also necessary to distinguish between installed solar power, RF power, and power delivered to the grid. Keeping this in mind, the following five metrics have been identified from the literature (e.g. Jaffe, 2020) as key determinants of space segment performance:

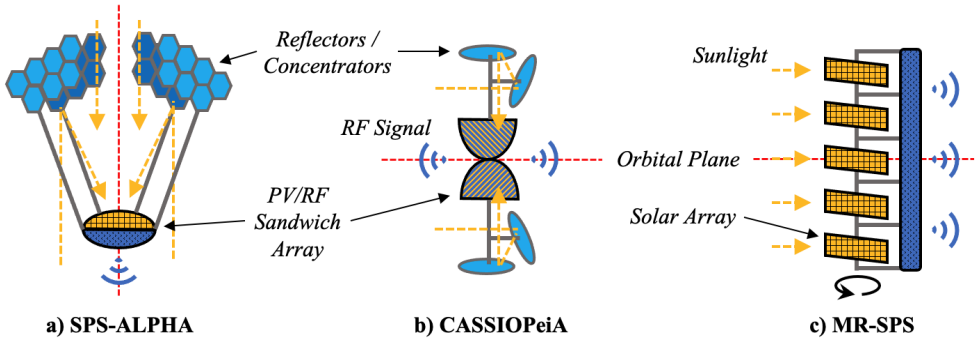


Figure 2: The satellite architectures of the three chosen concepts display significant differences in their approach on solar collection, conversion, and RF transmission. Own representation based on a) Mankins (2017), b) Cash (2019), and c) Hou and Li (2021).

Collect/transmit area-specific mass [g/m^2]:
This metric, also known as areal density, puts the surface areas of collection plus transmit apertures in relation with the overall weight of the space vehicle. Only collect and transmit areas are considered, as otherwise external reflector and concentrator structures with large surface areas but little weight could greatly distort this number. Generally, a decrease in area-specific weight is connected to a decrease in launch costs, manufacturing costs, and to a lesser extent an increase in transmit efficiency (Marshall et al., 2021). The mass is mostly driven by structural components such as any necessary trusses and booms, the solar and antenna modules, and the electronics. If the surface area is complex to determine, it is sometimes reported in square meters intercepted sunlight.

Mass-specific power [W/g]:
Mass-specific power measures the amount of mass that needs to be placed in orbit to achieve a given power level. Just as with the area-specific mass, weight is a key determinant of the economics of SBSP and here it is set in relation to the benefits gained in the form of power. For this metric, it is particularly important to distinguish between installed solar capacity and transmission capacity after additional conversion losses. We will refer to them as PV power and RF power, respectively. It is also essential to match PV power with PV weight and RF power with PV plus RF system weight to compare like with like and include all components necessary to achieve a certain power level. The power fed into the grid on the ground is unsuited for calculating specific power metrics used to compare space segments, as it is influenced by the parameters of the rectenna on earth. To allow for comparability, we will reference PV or RF power capacity of the space vehicle, depending on whether the focus is on the entire structure or just the solar part. Furthermore, many innovative prototypes for SBSP PV technologies have not fully integrated power beaming devices during their demonstrations. As a result, such experiments will provide an upper bound for mass-specific power levels given a certain technology and are hence still of relevance to our discussion. There might also be a trade-off between mass- and area-specific power, depending on whether launch capabilities are more constrained by weight or volume. However, the technologies discussed below suggest that weight might presently be a more determining factor, and hence we have chosen to put a larger focus on mass-specific power for the space segment. Together with the duty cycle, mass-specific energy can then be calculated.

Duty cycle [%]:
The duty cycle is measured as the share of one orbital rotation during which the SBSP can actively generate and transmit energy. Not only does this ratio depend on the structure and design of the satellite but also on the orbit to which the space segment is deployed. GEO is the orbit most commonly associated with SBSP concepts as it offers a theoretical duty cycle of 100%, depending on the satellite structure.

Conversion efficiency [%]:
Conversion efficiency for the space segment can either be measured from solar radiation to DC or DC-to-RF. For the solar components, cell efficiency also does not equal panel efficiency due to the wiring and other peripheral components. Efficiency is not only relevant to maximise the amount of power delivered to earth, but also in regards to heat management and consequently lifetime of the system. Generally speaking, the higher the efficiency, the less waste heat that needs dissipating will be produced.

Operating liftetime [a]:
The operating lifetime of the space asset is the amount of time the satellite can operate above a certain performance threshold. Of particular interest is the operating lifetime of key components, such as the solar cells. Here it is also seen as a proxy for the ability of the subsystems to deal with the harsh thermal and varying radiative environments present at all feasible orbits. Space maintenance capabilities also play a significant role in preserving the lifetime of the overall system. Nonetheless, the longer a component can function without the need for a replacement part, the fewer elements need

to be launched into space. Hence, the operating lifetime of components impacts the total mass that needs to be sent into orbit over the lifetime of the system.

To ensure that comparisons of metrics drawn from the literature are appropriate, some had to be recalculated or adjusted based on the disclosed data. For instance, a paper might report a mass-specific power metric for which only PV power but total sandwich module mass (including the antenna) was used. This number would not be suited for comparisons with metrics using only PV power and PV system mass or RF power and PV system plus antenna mass. Consequently, we had to recalculate some metrics based on the numbers reported with the demonstration results or adjust them based on data from other academic publications. To ensure metrics are adjusted following shared rules, we establish a common antenna weight premium that is added to purely PV-related technology options. This premium then allows us to determine an estimated mass for when these solar components are integrated into a sandwich module. The same concept applies for a structure premium, which accounts for supporting structures as well as deployment mechanisms and is added when individual sandwich modules are combined into a lightweight planar array. When applying these premia, it is assumed that the functional surface area stays constant. We also introduce common formulas to calculate mass-specific power and aerial density variants. These common formulas form the basis for any potential adjustments to ensure metrics include and exclude the same parameters when compared. The applicable formulas are introduced at the beginning of each subsection. A summary of all calculations is presented in section 2.7.

2.2. Photovoltaics

Solar power has a long history of being the energy source for projects in space, starting with the first solar-powered satellites Vanguard 1 and Sputnik 3 in 1958 (Andreev, 2018). Activity in space poses a number of demanding requirements to photovoltaic systems. They need to be lightweight, efficient and reliably operate for long periods of time in harsh conditions (Espinet-Gonzalez et al., 2019). The space environment is characterized by high and low energy particle radiation, large thermal cycles, high UV light exposure, and the possibility of collisions with space debris of any sort.

Having initially started with single-junction cells with only one junction to induce a flow of electrons, the most efficient cells today also used in space have multiple layers. These III-V multijunction cells are made from metal organic compounds of Group III and Group V elements, from which their name is derived. By correctly matching the subcell layers, thermal and transmission losses can be minimized (Philipps, Dimroth, & Bett, 2018) and as a result, power conversion efficiencies of the cells have increased to more than 50% over the last couple of years (Kalyuzhnyy et al., 2020). One of the techniques enabling the improved performance is photon recycling, where reflectors are placed to allow a second pass-through of unabsorbed photons through the photoactive region (Andreev, 2018). Specifically for multijunction cells, Bragg reflectors made of multiple semiconductor layers increase the absorption length of sunlight. Additionally, they also increase the radiation resistance. Consequently, great efficiencies, high reliability, a relatively high mass-specific power, and excellent radiation hardness have led to III-V multijunction cells commonly being used for satellites and space vehicles (Philipps et al., 2018).

In space, such cells are often used in conjunction with concentrators. By focusing light that would normally have impinged on a wider area on a smaller part of the cell, they are critical in keeping efficiencies high, reducing the amount of required cell material, and provide indirect radiation protection to the peripheral components (Andreev, 2018). Crucially for SBSP, they can also increase mass-specific power if lightweight materials are utilized and are necessary to achieve specific powers above 1 W/g (Warmann et al., 2020).

However, despite these benefits, problems remain. So far, space solar power was mostly designed around keeping an aperture running as efficiently and reliably as possible. With SBSP, this goal shifts towards the cost-efficient production of renewable energy for earth. Hence, factors determining the economics rise in importance. Weight is one of them, especially given that solar panels are typically the heaviest component of a satellite (Abdelal, Gad, & Abuelfoutouh, 2013). The weight of the cells themselves has already been cut with the introduction of thin-film variants, using thinned substrate (Law et al., 2006) and epitaxial lift-off technologies (Kayes, Zhang, Twist, Ding, & Higashi, 2014). Nonetheless, mass-specific power seems limited below 0.5 W/g (Gibb, 2018). Radiation shielding in the form of a cover glass placed on top of the cell is responsible for a significant part of the remaining weight (Espinet-Gonzalez et al., 2019). Hence, the cover glass necessary for today's most efficient GaAs multijunction cells is constraining mass-specific power and increasing area-specific mass. Nonetheless, ensuring an adequate useful lifespan by preventing degradation by radiation is also vital. This is especially the case for SBSP concepts in MEO, where the radiation environment is harsher than in GEO (Larson & Wertz, 1992). Different orbits and their implications are further discussed in section 2.6.

A second problem with current high-performance space cells is the reliance on geometric concentrators. As a result and as discussed in section 1.2, many modern SBSP designs include additional heavy concentrator structures comprised of booms and trusses to focus sunlight onto the solar cells without any intermittency or larger attitude adjustment maneuvers. Standard geometric concentrators also need to be pointed correctly to be effective, potentially requiring movable mechanisms constituting a potential point of failure. Other concepts also include additional on-chip concentrators (Cash, 2021a).

To measure the impact of these variations, we must first define common formulas for the mass-specific PV power p_{PV} and aerial density m_{PV}. Therefore, we designate these for-

mulas for the PV part of the satellite as follows:

$$p_{PV} = \frac{P_{PV}}{M_{solar}} \tag{1}$$

$$m_{PV} = \frac{M_{solar}}{A_{PV}} \tag{2}$$

$$M_{solar} = M_{SPG} + M_{reflectors} + M_{concentrators} \tag{3}$$

M_{SPG} denotes the mass of the solar power generating components and together with any reflector structures or concentrators thus forms the simplified total mass required to achieve the nominal solar power output of a system. Some brief calculations can now show the impact reflectors and concentrators can have on area-specific mass and mass-specific power in particular as well as provide us with some points of reference for our further analysis.

The SPS-ALPHA concept has a total orbital mass of 7,600 t with 3.4 GW of solar power installed across a surface area of about 2.2 km^2 (Mankins, 2021). The solar power generating units only make up about 5% of total mass (Mankins, Kaya, & Vasile, 2012) and hence theoretically achieve more than 8 W/g at 181 g/m^2. However, once the weight of concentrator and reflector structures is added in accordance with equation 3, this number nearly falls by a factor of ten to 0.95 W/g at about 1,600 g/m^2. This illustrates the desirability of concentrator PVs without bulky external structures, a concept we will investigate in sections 2.2.1 and 2.2.2.

Another example is the Chinese MR-SPS system. Due to its reliance on separate PV and antenna surfaces, it also requires great amounts of framing architecture. The total orbital mass exclusively attributable to the 2.4 GW of solar power generation is 2,000 t (Hou & Li, 2021). Despite a significantly lower density of 333 g/m^2, the resulting 1.20 W/g lie relatively close to the SPS-ALPHA concept due to the comparatively low power levels of MR-SPS. One reason for this might be the necessity of routing all the generated electricity through cables and slip rings to the antenna, which constraints the amount of power that can be installed to limit maximum voltage. Consequently, there is also no solar concentration employed. Mass-specific PV power would fall by half if the framing architecture were to be added to the equation.

Lastly, we will briefly look at the CASSIOPeiA concept, which reports a relatively light total orbital mass of 2,000 t and the collection of 11.3 GW of sunlight (Cash, 2021a). Accounting for optical losses of 20.1% due to multiple concentration steps and a PV cell efficiency of 39.4%, this results in 3.6 GW of PV capacity. Unfortunately, no further breakdown of satellite weight was available at the time of writing. Therefore, we can use the weight split of SPS-ALPHA as an approximation, with reflector and concentrators structures contributing 42% of total weight (Mankins, 2021). Based on the space segment architecture of CASSIOPeiA, where PV and atenna elements are jointly placed on the helical array, we assume a 50/50 split for the remainder. Consequently, specific PV power in accordance with equation 1 equals 2.51 W/g at 491 g/m^2. This can be considered very competitive compared to the other two concepts. However, the weight of external structures again decreases the metric by a factor of two to three, strengthening the case for a different approach to concentrator PVs for SBSP.

It should also be noted that neither of the three systems has an operational prototype and instead all fully rely on modelling to obtain their metrics. Hence, the numbers are theoretical and could decrease further in practice. In section 2.3 on module design, we also investigate similar dynamics when calculating performance indicators that include antennas and additional components. An overview of all metrics will then be provided in section 2.7. With the results above as a first point of reference for collect area-specific mass and mass-specific PV power, we can now investigate some technology alternatives that show promise in overcoming some of the aforementioned problems.

2.2.1. Integrated geometric concentrators

One possible solution to eliminate heavy supporting structures is to integrate concentrators directly onto the PV module. Such attachments could theoretically work in addition to any on-chip concentrators. One of the most successful designs replaces standard planar cell area with parabolic silver mirrors fabricated from ultralight carbon fibre reinforced polymer (CFRP) optics (Warmann et al., 2020). Each mirror would concentrate light onto the back of its neighbour, where a strip of PV cells including a cover glass is situated as pictured in figure 3. Additionally, a multilayer optical coating is applied to each mirror to aid with heat dissipation for cell cooling. As a result, cell temperatures can be kept below 100 °C despite the concentration. Heat management is particularly important for promising new cell materials as discussed in section 2.2.3.

It should further be noted that there is an inverse relationship between the level of concentration and the maximum acceptance angle under which sunlight can be gathered. While higher concentration increases specific power, it also constraints the angle from which sunlight can still be accepted. This trade-off has implications for overall module and system design as well as the power-optimal guidance of the structure.

Overall, on-module curved silver mirrors have greatly boosted specific power during experiments conducted by Warmann et al. (2020). The prototype module also included the RF components necessary for transmission but omitted any supporting structures required to form an array. We will return to the set-up including transmit antennas in section 2.3.1.

As part of the experiment, points of reference were established. Commercially available cover glass-interconnected cells (CICs) without any concentration displayed a mass-specific PV power of 0.54 W/g at 804 g/m^2 (Warmann et al., 2020). As a base for the concentrators, a mass-optimised multijunction space cell comprising a cover glass was used,

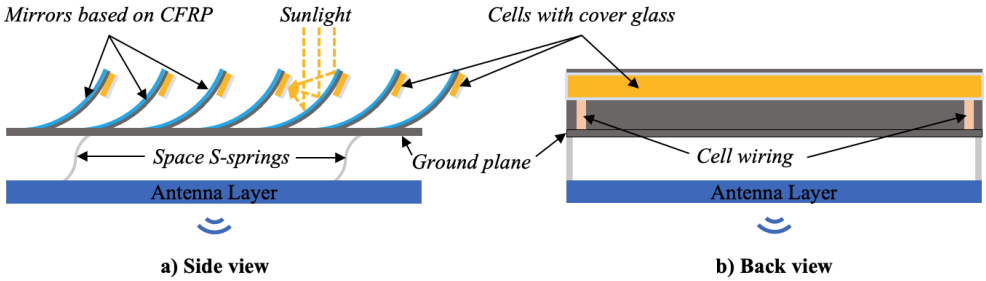


Figure 3: Visualisation of the ultralight concentrator concept by Caltech with parabolically curved mirrors concentrating the light onto strips of PV cells on the back of the neighbouring mirror. An integrated DC-to-RF and antenna layer has also been added. Own representation based on Warmann et al. (2020).

which on its own achieved 1.00 W/g at one sun and nearly half the area specific weight of 450 g/m^2. The addition of such a cell to a 7.5 sun concentrator structure, such as shown in figure 3, nearly quadrupled the result to 3.75 W/g at 116 g/m^2. Doubling the concentration factor to 15 suns and making the mirrors thinner led to another increase to 5.20 W/g at 83 g/m^2. By altering the coating on the concentrators, allowing the mirrors to be even thinner and hence lighter, a final increase of specific power to 5.90 W/g at only 67 g/m^2 was observed. This constitutes a significant improvement compared to the commercially available cells by around a factor of ten. Efficiencies were very similar at about 30% across all tested system compositions. The results also show significant improvements compared to the initial prototype tested two years earlier, which achieved only 0.23 W/g at 800 g/m2 (Kelzenberg et al., 2018). In comparison with some of the SBSP concepts, these measurements suggest that intelligently integrated concentrators could greatly boost the metrics critical for PV performance by increasing solar power levels while eliminating considerable amounts of mass.

Nonetheless, significant hurdles remain with the parabolic concentrator structures. Firstly, some of the ultrathin CFRP materials used in the demonstration are not yet commercially available, which renders mass-manufacturing capabilities uncertain (Warmann et al., 2020). Furthermore, volumetric structures on top of the module constrain overall system architecture due to PV capabilities being limited to only one side. Together with constraints on the acceptance angle under high-concentration scenarios, particularly accurate attitude management of the solar array would be required (Madonna, 2021). Lastly, they also complicate manufacturing processes overall and suffer from optical losses. Consequently, researchers have also been looking at other integrated concentrator alternatives which allow for planar PV structures. Depending on the SBSP system design, the benefits of maintaining a planar structure to enable dual-sided modules can be substantial and are further discussed in section 2.3.

2.2.2. Luminescent solar concentrators

To maintain a flat-plate architecture while keeping specific power high and area-specific weight low, a different approach is required. Developing photonic devices with characteristics at the nano-scale opens up possibilities for new non-geometric concentrators by taking advantage of elemental interactions between light and matter. Luminescent solar concentrators (LSCs) make use of luminophores in the form of quantum dots, which emit photons when excited by short-wave light or ionising radiation (Needell et al., 2017). Quantum dots are nano-scale material structures made from semiconductors whose extremely small scale leads to changes in their optical and electronic properties based on quantum mechanics. These can then be manipulated for the purpose of concentrating solar irradiation on a PV cell by embedding them into an optical waveguide. The resulting architecture is displayed in figure 4. When hit by radiation passing through the layers of the array, the luminophores then emit light which is directed towards a PV micro-cell by the waveguide. Through selectively picking the semiconductor materials of which the quantum dot is composed, it is possible to influence the bandwidth of wavelengths in which light is emitted. This bandwidth can then be optimized for the chosen solar cell to maximize efficiency, for example by down-converting blue or near-UV light to red (Madonna, 2021).

While photovoltaic systems utilizing standard geometric concentrators, such as described in section 2.2.1, are particularly bad at capturing diffuse sunlight, LSCs can absorb direct and diffuse sunlight (Needell et al., 2018). Therefore, quantum dots can not only eliminate the need for concentrator structures but also boost efficiency further by capturing more of the light available. These benefits have also been observable in device modelling simulations utilizing LSCs, after first tests in 2018 indicated that quantum dots could notably improve conversion efficiencies (Needell et al., 2018). Without bandwidth optimization of the light emitted by the luminophore, mass-specific power levels of above 1 W/g at 140 g/m^2 and conversion efficiencies of about 12% have been achieved (Needell et al., 2019). By using spectrally-optimized quantum dots, specific power values were nearly doubled to 1.84 W/g at the same area-specific mass and efficiencies improved to close to 20% . Other simulations sug-

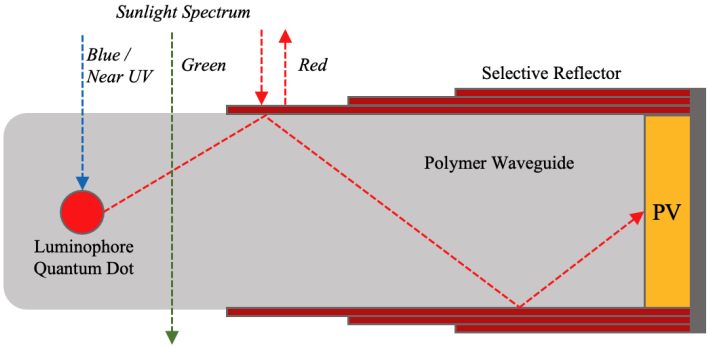


Figure 4: Luminescent solar concentrators employ a quantum dot made from a luminophore which emits light of specific wavelenghts when excited by radiation. This light is then directed towards a solar cell by the waveguide. In this example, near UV light passing through a selective reflector is converted to red light by the quantum dot before being transmitted to the PV cell. Own representation based on Madonna (2021).

gest that, under optimal conditions, power conversion efficiencies of around 30% are well within reach (Needell et al., 2017). In first tests, LSC cells have also exhibited higher radiation tolerances due to a reduction in their effective cross section (Hu et al., 2021).

Therefore, quantum dot cells appear to be a promising option for delivering radiation resistant flat-plate concentrator PV modules. LSC technology can also be integrated into flexible PV sheets (Needell et al., 2019). This characteristic is greatly beneficial for space solar module designs that rely on folding or coiling to keep volumes small during launch. A material of particular interest for these types of flexible quantum dot cells has been perovskite. Despite their relatively early stage of development, flexible LSC perovskite cells have achieved efficiencies of more than 12% and displayed greater mechanical endurance than standard thin-film PV cells (Hu et al., 2021). A research grade device of this kind is set to fly for tests in 2022 (Madonna, 2021). Additional benefits of perovskite are discussed in more detail in the following section.

However, while LSC modules are certainly promising, it should be noted that simulations did not include any connecting or transmission structures, which naturally decreases weight. We will address this in the conclusion of this chapter by utilizing the calculated weight premia. Additionally, two limitations that were frequently encountered are the absorption of the light by the luminophore as well as poor trapping of the emitted light by the waveguide (Needell et al., 2018). Both of these result in less solar energy reaching the PV cell and have to be investigated going forward. Lastly, their exact behaviour under thermal and radiation stress has not been fully explored yet. Future research efforts will have to address these obstacles to make LSC cells ready for SBSP systems, given the promise they are showing in attaining high mass-specific power levels while maintaining a planar module geometry.

2.2.3. Perovskite cells

Researchers are continuously exploring new material compositions with beneficial properties for the use in solar cells. One of the materials that has gathered particular attention over the last couple of years is perovskite. Solar cells based on perovskite display the ability to withstand high-energy proton doses that exceed the levels basic silicon cells can endure by almost three orders of magnitude (Lang et al., 2016). But not only is their radiation resistance higher, their self-healing properties mean that photocurrent and photovoltaic performance of the cell also start to recover over time. This process based on thermal annealing even starts when performance has dropped to as low as 2% after high radiation exposure (Madonna, 2021).

The strong inherent radiation tolerance and high damage threshold would allow for the removal of the protective cover glass. Lightweight flexible perovskite solar cells manufactured this way have been tested under radiation conditions that equal several years of exposure in space, confirming their resilience and great potential for SBSP applications (Malinkiewicz, Imaizumi, Sapkota, Ohshima, & Öz, 2020). First prototype cells based on perovskite compounds have been able to achieve exceptionally high specific power values of 29 W/g at an efficiency of 15.2% (Kang et al., 2019). The aerial density m_{PV} was measured at a mere 4 g/m^2, in part due to the lack of any heavy glass. While it should be noted that this was for a single prototype cell without any peripheral structures and electronics, the performance increase is still significant and would even remain very competitive with mass increases by a factor of ten or more. However, at about 20%, power conversion efficiencies still notably lag behind those of standard multijunction cells (Kim et al., 2017).

Nonetheless, perovskite cells also display problematic characteristics. Thermal degradation already starts to occur at 85 °C (Kim et al., 2017), with performance deteriorating in less than 24 hours to the point of inoperability as a result (Miyazawa et al., 2018). This is particularly problematic as the solar panels of satellites are known to be one of the

parts most exposed to severe thermal cycles, reaching temperatures of up to 125 °C (Pisacane, 2005). While progress has been made in composing cells that are more stable and can manage elevated temperatures for extended periods of time, the conversion efficiencies of these variants have so far remained very low at under 5% (Miyazawa et al., 2018).

In conclusion, perovskite solar cells are still relatively early in their development compared to modules based on other materials. Still, more research should and is being conducted on how to keep the coveted characteristics while overcoming some of the shortcomings, given the great promise these first forays have shown.

2.2.4. Nanowires

Besides for the purpose of solar concentration, operating at the nano-scale can also be exploited to achieve greater radiation resistance. Having inherently radiation-resistant materials that require minimal additional protection in the form of heavy cover glass can potentially boost specific power ratios more than simple efficiency gains (Gibb, 2018). Based on this idea, material or structural compositions that trade efficiency for radiation resistance in order to eliminate the cover glass and extend the operating lifetime are of particular interest.

One technology that follows this approach are nanowire PV cells. Nanowires are one-dimensional nanostructures used for electrical transport (Cui, Duan, Hu, & Lieber, 2000). Therefore, they effectuate the same purpose as normal wires in standard cells but at a much smaller scale. Together they form an array of high-aspect-ratio semiconductor structures with particular dimensions to enhance light absorption and radiation resistance (Barrigón, Heurlin, Bi, Monemar, & Samuelson, 2019). Simulations have shown the great potential of this technology to provide efficient, lightweight, and radiation-tolerant power generation units in space (Espinet-Gonzalez et al., 2019). These findings could then be confirmed by experiments, in which nanowire cells displayed a damage threshold that was 10 to 40 times higher compared to standard planar control cells. Upon closer inspection, this benefit seems to result from a reduction in effective cross section susceptible to defect production, much like with LSCs. In other words, by using wires at the nano-scale arranged in a particular fashion and embedded in the right materials, these cells present a substantially smaller area where impingement by incoming high-energy particles would lead to degradation. They also benefit from the fact that they are otherwise made of the same materials as standard III-V multijunction space cells.

Tests under space conditions were further able to corroborate these findings, with efficiencies ranging from 15% to about 18% and improving over time (Espinet-Gonzalez et al., 2020). Researchers also expect potential for optimization with regards to the array geometry, which could further boost efficiencies and radiation tolerances. Nonetheless, these numbers are the evidence of a for now necessary trade-off between efficiency and radiation tolerance compared to standard planar arrays. Lastly, nanowire arrays can also be produced in the lightweight flexible sheet form conducive to packaging and deployment in space (Cavalli, Dijkstra, Haverkort, & Bakkers, 2018). Overall, they provide the possibility of extending operating lifetimes while eliminating the cover glass for SBSP applications, in turn increasing specific power through a reduction in specific weight.

While the exact magnitude of these gains will have to be validated through additional prototype testing, some brief calculations can show the potential. We assume that nanowire technology will offer the possibility of completely eliminating the cover glass used for a specific cell. In reality, this will likely only be possible in tandem with some of the other technologies introduced in this section. However, for simplicity, we will fully attribute the potential benefits to nanowires for the purpose of establishing an upper bound. Expecting a convergence of efficiency levels over the long term and knowing that the cover glass can account for about two thirds of total PV mass (Xu, Li, Tan, Peters, & Yang, 2018), we use the commercially available CIC from Warmann et al. (2020) to recalculate mass-specific PV power with the weight reduced by two thirds. The result of 1.56 W/g suggests that nanowire cells bear the theoretical potential of tripling p_{PV} from initially 0.54 W/g by decreasing aerial density of the CIC from 804 g/m^2 to 278 g/m^2.

2.3. Module design

While reliable and lightweight PV is critical for the functionality of the space segment, it is at least of equal relevance how the solar elements are integrated into modules with the other satellite components, notably the transmission devices. The importance of such modularity is further addressed in chapter 4.2. In this section, we introduce various design options and their effect on weight- and power-related metrics. Possible satellite structures and resulting deployment dynamics will be discussed in the subsequent section.

Including the RF infrastructure of the space segment into our analysis requires us to expand our equations from section 2.2. Adding the mass and surface area of transmission hardware components as well as considering the RF power output of the satellite yields the following formulas.

$$p_{RF} = \frac{P_{RF}}{M_{solar} + M_{RF}} \tag{4}$$

$$m_{RF} = \frac{M_{solar} + M_{RF}}{A_{PV/RF}} \tag{5}$$

$$P_{RF} = P_{PV} * \eta_{DCRF} * \eta_{antenna} \tag{6}$$

$$M_{RF} = M_{DCRF} + M_{antenna} \tag{7}$$

The index $DCRF$ denotes the infrastructure necessary to convert the DC power generated by the solar array into RF power, which is then transmitted by the antenna. In some

publications η_{DCRF} also already includes $\eta_{antenna}$. Additionally, $A_{PV/RF}$ varies depending on the satellite design. For sandwich modules, which have a shared collect and transmit structure, equation 8 applies.

$$A_{PV/RF} = A_{PV} = A_{RF} \tag{8}$$

For separate solar and transmission structures, such as with the MR-SPS concept, $A_{PV/RF}$ is defined by equation 9.

$$A_{PV/RF} = A_{PV} + A_{RF} \tag{9}$$

The different metrics reported in Warmann et al. (2020) for modules with and without the RF component also allow us to determine a general weight premium for adding transmission infrastructure to a lightweight solar module. The magnitude of this change is therefore measured at 29 g/m^2. This premium can then be used heuristically to turn standard PV modules into integrated sandwich modules and vice versa. We have therefore established a common baseline for calculating the metrics selected to compare different satellite module designs.

2.3.1. Sandwich modules

As discussed in section 1.1, most historical SBSP designs relied on separate solar collection and power transmission surfaces. Even today, the Chinese MR-SPS system still relies on an external antenna connected to the solar array via rotating joints (Hou & Li, 2021). However, there are two major problems arising from such an approach. First, the 6-km^2 large solar array requires extensive wiring to route the electricity to the point of transmission. Second, the for accurate targeting necessary independent rotation of both surfaces requires something like a slip ring mechanism, which will be under considerable thermal stress and constitutes a single point of failure (Jaffe, 2020). Therefore, almost all modern designs envision the use of so-called sandwich modules to drastically reduce overall weight, in line with the trends we have seen for solar components. These integrated modules, which allow for the elimination of a majority of the wiring and supporting structures during the DC-to-RF conversion process while also introducing more redundancy to the system, can be used in conjunction with lighter PV and transmission technology to achieve higher specific power levels. The California Institute of Technology is one of the leaders in the area of planar sandwich modules, thanks to its extensive research efforts as we will see below.

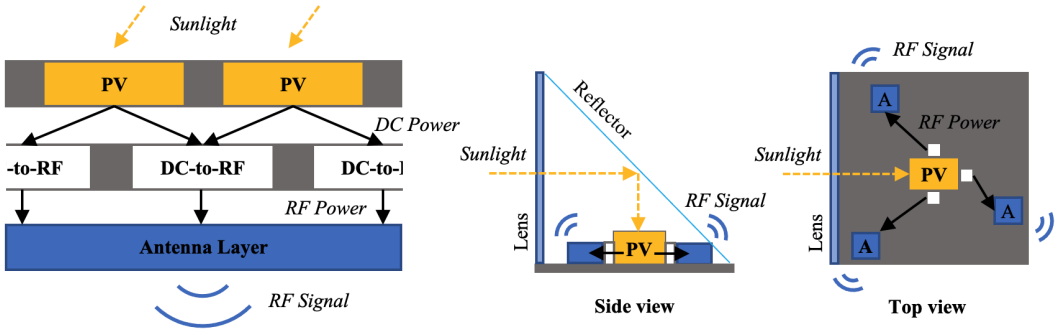
Generally speaking and as the name suggests, sandwich modules consist of multiple layers, each of which dedicated to one of the key functions of a SBSP satellite (S. A. Hajimiri, Atwater, Pellegrino, Abiri, & Bohn, 2021). Hence, there are at least three layers as shown on the left in figure 5. One for PV-to-DC, another for generation of the microwave signal or DC-to-RF, and a third for the transmission of the signal. Additional layers might be required, e.g. for beam control. Together, they form a module that acts as a solar-powered space-borne transmit antenna, sometimes called a spacetenna (Jaffe, 2020). As a result, routing distances are minimized, heavy cables and many supporting structures eliminated, and the spacetenna is deconstructed into many identical individual parts (Madonna, 2018). Such modularity also has significant implications for manufacturing and maintenance as discussed in section 4.2. For some applications it can be advantageous to rearrange the layers of the sandwich module from a flat tile into an L-shape (Jaffe, 2013) as shown on the right in figure 5. However, this only changes the geometry due to added optics but not the underlying functionality of the panel as the elements still share a common substrate and are therefore nonetheless considered a sandwich module.

The idea of sandwich architecture was first introduced by Owen Maynard as part of early SBSP studies by NASA and the Department of Energy (Maynard & Blick, 1980). Since then, first prototypes were introduced by Japanese researchers around the turn of the millennium (Matsumoto, 2002) and later further developed at Kobe University (Etani, Iwashita, & Kaya, 2011). Unfortunately, metrics relevant for specific power or aerial density calculation were reported for neither of the two. Another prototype was built at the U.S. Naval Research Lab (NRL) and tested in realistic space conditions with varying illumination (Jaffe, 2013). However, mass-specific power remained rather subdued due to a very high aerial density of 21.9 kg/m^2, resulting in merely around 0.005 W/g.

Significant progress has been made since, as evident by the sandwich modules developed by Warmann et al. (2020) and introduced in section 2.2.1. Once the conversion infrastructure and an antenna were added to the PV section, as displayed in figure 3, even a mass-optimised multijunction cell without any concentration still achieved a mass-specific RF power p_{RF} of 0.67 W/g at 476 g/m^2. With concentrators and the lightweight redesign, the metric went up to 2.87 W/g at 96 g/m^2 for these kind of sandwich modules. While the specific RF power level is less than half that of p_{PV}, it can still be considered as quite competitive. However, it should also be noted that many prototypes have only integrated the elements indispensable to the solar-to-RF process and are missing phase shifting equipment or sometimes even a functioning antenna. Therefore, slight weight increases should be expected for a SBSP-ready sandwich module.

In comparison, the Chinese system with its separated generation and transmission surfaces, only achieves 0.27 W/g in mass-specific RF power when including PV and antenna systems but excluding supporting structures (Hou & Li, 2021). These calculations suggest that combining both surface areas via sandwich modules could significantly increase the relevant performance indicators of the space segment.

Nonetheless, the way in which the module layers are integrated and the weight of the antenna components also matters. The SPS-ALPHA concept envisions hexagonal frames to combine the functional layers (Mankins et al., 2012). These frames add a lot of mass to an otherwise relatively light sandwich design. Consequently, mass-specific RF power falls to 0.33 W/g at a comparatively heavy 3,222 g/m^2.



a) Standard sandwich module

b) L-shaped sandwich module from CASSIOPeiA

Figure 5: The integrated layers and components of two variants of sandwich modules. The strategically placed antennas in b) allow for 360-degrees beam steering on the horizontal plane. Own representation based on a) Madonna (2018) and b) Cash (2017).

For CASSIOPeiA, we still have to include the helical array structure to which the modules are affixed due to a lack of granular data. Nonetheless, it achieves a notable p_{RF} of 1.51 W/g at only 692 g/m^2, even including these structures. The fact that CASSIOPeiA therefore outperforms the MR-SPS by a factor of three further suggests that the benefits of well-integrated transmission and collection surfaces are substantial.

An improved version of the sandwich prototype proposed by Jaffe (2013) has also recently been sent for in-space testing with the help of the X-37B orbital test vehicle (NRL, 2020). During the NRL's Photovoltaic Radiofrequency Antenna Module Flight Experiment (PRAM-FX), the test vehicle ascended to LEO to observe conversion efficiencies under space conditions. The modules were also occasionally heated to demonstrate the thermal conditions in GEO. Preliminary results suggest DC-to-RF efficiencies of 37.1% and PV efficiencies of 22.6%, resulting in a rather low overall module efficiency of 8.4% (Rodenbeck et al., 2021). This is symbolic for some of the issues that remain with such tiles.

Space segment design relying on sandwich modules also has inherent shortcomings. Integrating all functionalities and consequently the materials into a single tile can prove difficult while maintaining high efficiencies and pose problems for heat dissipation (Jaffe, 2020). Hence, the efficiency of each layer becomes even more important to minimise thermal losses. Another issue is the correct alignment of the individual tiles. The structure resulting from these individual modules is susceptible to applied forces such as microgravity and therefore requires in-situm surface shape measurement to maintain beam control (Madonna, 2018). This issue of array management is discussed in section 2.5.

Furthermore, the orientation of the sandwich structure during its orbit to ensure continuous operation can pose problems. Seeing how solar and antenna surfaces can no longer be rotated independently of each other, simultaneous alignment towards the sun and the ground station becomes impossible for some orbital positions. For instance, the satellite would not be able to transmit around local midnight, as the antenna surface would point away from earth if the solar side remains sun-referenced. This problem can either be addressed at the module design or the satellite structure level. For structural solutions, rear reflectors or mirrors combined with concentrators, such as the ones used in the SPS-Alpha (Mankins, 2021) or CASSIOPeiA (Way & Lamyman, 2021b) concepts, would trade additional weight for a continuous duty cycle. However, focussing light on the sandwich tiles also bears the risk of increasing any thermal challenges. In a more innovative approach, the CASSIOPeiA's helical arrangement could pose another potential solution. This structural concept is briefly examined in section 2.4.2. Another alternative on the design side are dual-sided sandwich modules, which we investigate next.

2.3.2. Dual-sided modules

One idea that has been investigated to overcome duty cycle limitations of planar structures comprised of sandwich modules are dual-sided tiles (Marshall, Goel, & Pellegrino, 2020). In general, this would result in the integration of either a second RF surface, a second PV surface, or both, allowing for increases of the duty cycle of up to 50%. These different configurations are displayed in figure 6.

In order for the supplementary layers to not block solar or RF functionalities on any side, they either have to be optically transparent RF or RF-transparent PV. Due to the transparency, adding both only results in greater module weight without any meaningful expansion of the duty cycle. With one dual-sided capability, it is already possible to optimise for the attitude of the single-sided element without compromising the functionality of the other layers. However, optically transparent antennas and particularly RF-transparent PV is technologically very challenging (Madonna, 2018). Hence, development so far has mostly focused on sandwich modules with an additional optically transparent RF layer, which also promises to be lighter than the alternative of transparent PV. Dual-sided RF furthermore benefits from shared DC-to-RF

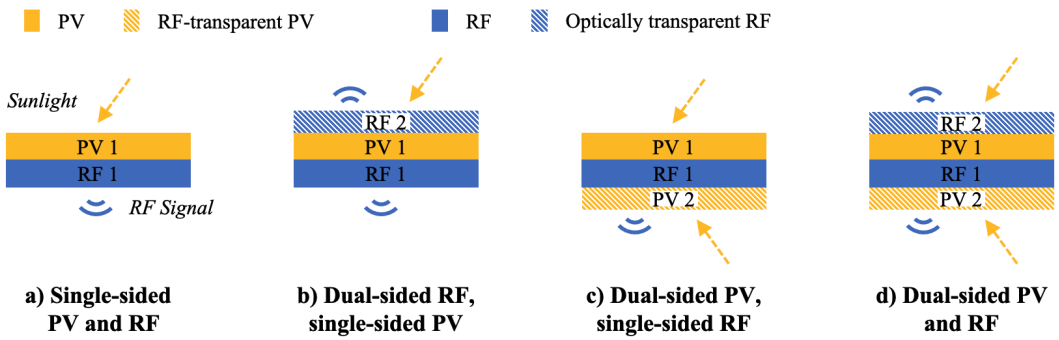


Figure 6: Sequential addition of transparent layers showcasing different possible configurations for dual-sided sandwich modules. The addition of one transparent layer as in b) or c) is enough to extend the duty cycle by up to 50% for planar arrays without rear reflectors. Own representation based on Marshall et al. (2020).

infrastructure.

Transparent antennas are not a completely new concept. Already in 2009, American researchers have experimented with mesh antennas placed on the solar cells of small satellites, achieving transparency levels of 93% (Turpin & Baktur, 2009). Overall, a meta study in 2020 was able to confirm that their performance had reached close to parity with non-transparent kinds (Silva, Valenta, & Durgin, 2020). Similar concepts have also been investigated by Chinese scientists (Qiu et al., 2021). Therefore, these findings suggest that adding an optically transparent RF layer appears to be a promising way to maximise the duty cycle of planar sandwich arrays without greatly increasing module weight.

However, even with dual-sided tiles in place, precise attitude control is still essential. This optimisation problem is also called power-optimal guidance and is based on a trade-off between a position that maximises power collection and a position that maximises power transmission while accounting for the permissible squint angles of both surfaces (Marshall et al., 2020). All these parameters also vary with the chosen orbit, with GEO displaying the greatest benefits from introducing dual-sided variants.

While no data on dual-sided RF module prototypes has been reported at the time of writing, we can once again use the CIC and flat-plate cell from Warmann et al. (2020) to calculate some approximations. Essential for this heuristic is the determination of a multiplier for the standard one-sided RF layer weight premium established at the beginning of this section. The fact that state-of-the-art optically transparent antennas are of a mesh-like structure and can share significant parts of the DC-to-RF conversion infrastructure of the other antenna suggests that the multiplier should be well below two. For the purpose of our approximation, we have chosen 1.3 as a suitable factor. Once these assumptions are applied to the cells' parameters, we obtain 0.28 W/g for the CIC and 0.51 W/g for the flat-plate variant. Both constitute a decrease of about 20% when compared to their standard sandwich configuration. This gives us a rough order of magnitude for the benefits an extended duty cycle would have to realize to outweigh the increase in weight. To fully capture this dynamic, mass-specific energy might be a better metric for comparing single- and dual-sided sandwich modules. Nonetheless, additional research beyond models is needed to investigate whether overall the mass, cost, and complexity added by introducing additional layers ultimately overshadow any system efficiency gains (Marshall et al., 2020).

2.4. Satellite structure and deployment

In this section, we explore different satellite structures into which the individual modules can be arranged. Some of our considerations from the previous sections, especially those on concentrators, will be of great relevance to the overall structure. In particular, we will explore the impact of the addition of trusses, booms, and any other structural elements necessary to form the space segment on its weight- and power-related metrics.

Our primary focus when it comes to SBSP satellite structures rests with planar and helical shapes, the latter addressing some of the shortcomings of flat-plate architectures. Lastly, we also consider any stowage and deployment mechanisms required to place the body in orbit, as these also impact satellite design and the selected performance metrics. Therefore, compact packaging and smooth deployment while adding as little weight as possible are essential.

Consequently, we will also have to add new variables for supporting structures to our equations to calculate overall satellite array performance. Mass-specific power and aerial density for the entire space segment will therefore be calculated in accordance with equations 10 and 11.

$$p_{satellite} = \frac{P_{RF}}{M_{solar} + M_{RF} + M_{structures} + M_{auxiliary}} \quad (10)$$

$$m_{satellite} = \frac{M_{solar} + M_{RF} + M_{structures} + M_{auxiliary}}{A_{PV/RF}} \quad (11)$$

For $A_{PV/RF}$, equations 8 and 9still apply. The mass of auxiliary systems such as propulsion is represented by $M_{auxiliary}$. However, in most concepts they account for less than 1% of total mass, limiting their impact on $p_{satellite}$ and $m_{satellite}$.

These formulas now allow us to determine the effect of including structures to form a stowable and deployable array out of individual modules.

2.4.1. Planar arrays

Planar arrays combine a large number of sandwich tiles or modules into one common plane, which acts as a solar power collector and converter as well as transmission antenna in space. One envisioned approach for this is displayed in figure 7. A tile is seen as the basic unit of functionality of which many will be arranged and connected into a strip with 1 m in width and 60 m in length (Pellegrino et al., 2020). Attached to deployable booms, these then form a PV and RF space segment measuring 60 m × 60 m. Multiple of these space segments combined could make up a SBSP power station. Therefore, planar arrays are easily scalable by adding to the number of space segments. Consequently, sequentially increasing the size of any smaller prototype in space until commercial scale is achieved should be possible without greater issues.

Once again, the use of dual-sided tiles as discussed in section 2.3.2 would be essential to maximising the duty cycle of a planar space segment. So far, increases of up to 50% are expected compared to single-sided modules (Marshall et al., 2020). However, some intermittency would remain, requiring utility-scale storage on the ground to achieve constant power supply. Therefore, the overall trade-off for a planar satellite would be between significant advantages in terms of modular structure, scaling, and weight versus some remaining intermittency. However, large-scale storage is also critical for other more prominent renewable technologies such as wind and terrestrial solar. Hence, this is not an issue specific to SBSP and contrary to the other technologies, the intermittency associated with space solar would be very consistent and predictable over the entire lifetime of the satellite.

For large planar structures to be viable, a mechanism needs to be in place to allow for compact packaging during space launch and transport as well as the subsequent deployment. One approach would see the space segment being z-folded and rolled prior to launch, inspired by specific origami techniques called *kirigami* (Gdoutos, Leclerc, Royer, Türk, & Pellegrino, 2019). Following such a procedure, the nominally 60 m × 60 m payload could be transformed into a cylindrical package of 2.2 m diameter and 1.8 m height (Madonna, 2021). This is based on extrapolations from smaller scale lab studies using a 1.7 m × 1.7 m prototype (Gdoutos et al., 2020).

Once the orbit is reached, deployment is initiated by first uncoiling the cylinder and then unfolding the different strips (Pedivellano, Gdoutos, & Pellegrino, 2020). This is achieved by releasing strategically placed external constraints and subsequently letting the array self-deploy, using the elastic energy stored during packaging. Such an approach is also called strain-energy-deployed. Nonetheless, a 60 m × 60 m structure would have many folds susceptible to wrong or incomplete deployment. Therefore, the sequence in which the constraints are released needs to be optimised and controlled to minimise chances of failure. First trials suggest that a stepwise deployment from the inner-most to the outer most strips of the space segment could be a viable option (Pedivellano et al., 2020).

Tests of space segments for such a satellite structure have already been able to provide some insights into the metrics, with an initial design in 2016 setting a target of 100 g/m^2. A prototype with the dimensions of 1.7 m × 1.7 m managed to achieve 150 g/m^2 in 2018. At a share of 42%, the two longerons connected to each strip contributed the most to the system's overall mass (Gdoutos et al., 2020). The 50-µm thick functional material followed closely behind at 41%. Functional area also makes up about 75% of the total area of the space segment. The result was an areal density of 136 g/m^2, which is comparable to the levels discussed in section 2.3 on module design. Additionally, as the structure is scaled up to its commercial proportions of 60 m × 60 m, the weight share of supporting structures nearly halves to 24%, bringing area-specific mass down to 126 g/m^2. Such numbers would be quite comparable to some of the other concepts such as SPS-ALPHA, where supporting structures at scale are responsible for a similar share of total mass (Mankins et al., 2012). However, SPS-ALPHA is still much heavier at an absolute level per module, suggesting room for improvement.

For the extrapolation of the commercial array, an aerial density of 100 g/m^2 for the functional material was assumed, in line with what has been achieved in past demonstrations of sandwich modules (e.g. Warmann et al., 2020). Based on the numbers reported by Gdoutos et al. (2020), we can assume that additional structures to form the space segment as well as for deployment and packaging would increase the aerial density of high-performance ultralight sandwich modules on average by 32 g/m^2. This structural weight premium can now be used to heuristically calculate performance metrics for an array containing any type of sandwich module. For instance, adding structures to the flat-plate multijunction tiles from section 2.2.1 would lower mass-specific RF power only slightly from 0.67 W/g to 0.62 W/g. The impact is greater with lighter modules such as the 15-suns concentration variant, where specific power would decrease by about a fourth to 2.15 W/g. Such a value for $p_{satellite}$ would still be significantly higher than CASSIOPeiA's mass-specific RF power for the entire satellite of 1.51 W/g. This suggests that the benefits of eliminating external reflector and concentration structures might be substantial. Due to their heavy components and architecture, SPS-ALPHA and MR-SPS only achieve 0.32 W/g and 0.16 W/g, respectively.

There are also efforts to eliminate even more connectors and hence mass by having the elements of the space segment fly in formation. For instance, the private company Solaren has patented one specific approach (Rogers & Spirnak, 2005). The idea is that power plants comprised of multiple space segments orbiting in formation would be even easier to scale and require less assembly. However, the power-optimal guidance problem becomes greater as the number of independent elements increases. First studies indicate that the elements of the constellation would also have to perform peri-

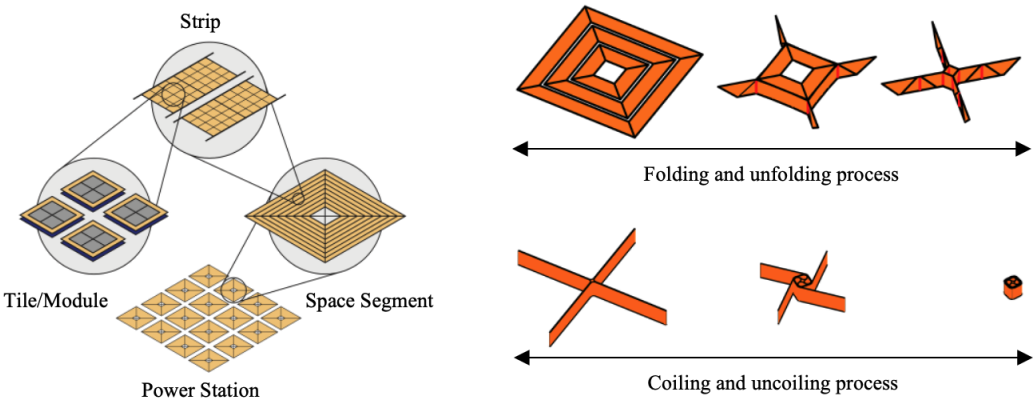


Figure 7: The left side is a visualisation of how individual tiles or modules form space segments which can be combined into a space power station. The right side shows the deployment mechanism for the space segments as a sequential combination of folding and coiling. Adaptation based on Gdoutos, Truong, Pedivellano, Royer, and Pellegrino (2020) and Madonna (2021).

odic orbits relative to each other in order to not obstruct other space segments (Goel, Lee, & Pellegrino, 2017). Algorithms to optimize such relative movements within the constellation are also being developed (Goel, Chung, & Pellegrino, 2017).

2.4.2. Helical arrays

As briefly mentioned before, the CASSIOPeiA concept is pursuing a very different solution to avoid the power-optimal guidance problem. By maintaining the underlying concept of sandwich modules but arranging them along a double-helical structure around a central axis instead of in a planar fashion, a continuous duty cycle in GEO could be achieved without sudden attitude adjustments (Cash, 2021a). As a result of the change in satellite design, L-shaped sandwich module variants as shown in figure 5 are employed, with PV and RF components placed on a common substrate on the horizontal plane. The other branch of the L would mostly consist of smaller struts to connect to the adjacent horizontal layers as well as lenses and secondary reflectors. The antennas would be spaced on the substrate in consideration of the employed wavelength.

As introduced in section 1.2, the helix would be complemented by two solid state-symmetrical reflectors on each end (Cash, 2019). The dual primary reflectors above and below the orbital plane would be sun-referenced and focus the light onto the helix. By collimating the sunlight, they would ensure uniform lighting of the sandwich array. Additional secondary reflectors integrated into the sandwich modules would then re-collimate the rays onto the PV chips. While the current module design for CASSIOPeiA is planning on using standard III-V multijunction space cells with on-chip concentration, the technologies discussed in section 2.2 could theoretically also be applied to further optimise for weight, power, efficiencies, and radiation resistance. Given that the aperture of a helix is constant from every angle, the structure itself can furthermore intercept an equal amount of sunlight from every side. However, since the reflectors providing the first step of concentration are static, the satellite has one preferred orientation towards the sun (Cash, 2021b). Therefore, the structure will have to rotate while orbiting to ensure the side which maximises light collection is always directed towards the sun.

So far, we seem to be facing a similar problem with the helical array as with the planar array. With the orientation relative to the sun remaining constant, the orientation towards the receiving station on earth is perpetually changing. However, superior flexibility of the antenna function integrated into the helical array aims to tackle this issue. Each of the L-shaped sandwich modules would be comprised of three flexible dipole antennas forming one antenna element which can be associated with one or multiple PV devices on a common substrate (Cash, 2021b). By strategically placing these three flexible dipole antennas on each sandwich module, omnidirectionality for the microwave beam can be achieved. The projected area of the helix remaining constant from every angle also means that the surface area of the transmitting aperture facing the rectenna will technically remain constant despite the rotation (Cash, 2017). All antennas would then be connected to a system-wide synchronized timing reference source to manage the 360-degrees steering capabilities of the antenna array through phase shifting of the individual signals. This is discussed in more detail in the following section. Overall, the structure can therefore slowly gyrate to maintain optimal orientation towards the sun while continuously transmitting to a fixed point on earth due to its collective 360-degrees steering window. As we have seen, the resulting continuously available mass-specific RF power of 1.51 W/g can be considered very competitive.

Despite these clear theoretical advantages, which have yet to be confirmed through a prototype, some drawbacks remain. Due to the increased complexity of the structure, it is not as easily scalable as planar arrays. To increase the power capacity in space, an entire new satellite would have to be launched each time instead of just adding parts to the existing one. Due to the use of sandwich modules, a certain portion of modularity is still maintained, which will be beneficial for

manufacturing and repairs. Similarly to the deployment of the planar variants, the nominally 1.7-km tall core structure of the helical array would be collapsible and enable automatic unfolding with the help of springs and releases (Cash, 2021b). The targeted dimensions for the packaged helix have not been disclosed yet. However, autonomous space robots are anticipated to conduct some of the in-orbit assembly. This issue is briefly discussed in section 4.2.

2.5. Array management

Essentially, array management fills a critical role as enabler of the modular sandwich structure. The benefits of modularity for SBSP systems inevitably result in the antenna being broken up into multiple smaller transmission devices. In this section we discuss how an array comprised of a large number of individual antenna elements can still act like one big antenna by exploiting some of the basic characteristics of waves. While such an approach may come with a number of advantages compared to a singular big antenna, it also poses some technical hurdles. This is particularly true for structures as large as the ones required for SBSP. As the coordination between the modules requires some form of connection, often in the form of cabling, it would add complexity and weight to the space segment. However, there are some new approaches that aim to overcome this. Nonetheless, the underlying concept is rather intuitive and essentially does not differ between planar and helical arrays.

2.5.1. Phased arrays

Generally, forming a continuous microwave beam carrying large amounts of energy over long distances is challenging. As a result, microwave power sources exceeding 10 kW are typically pulsed as devices capable of continuous transmission are rare, expensive, and difficult to maintain (Rodenbeck et al., 2021). All these characteristics make them ill-suited for SBSP. Instead, multiple smaller apertures arranged on a common surface can send out beams with lower power in a concerted fashion, which then reach their target as if they were a single high-power signal.

However, this set-up requires extensive coordination between the individual elements regarding targeting, amplitude, and phase alignment. Synchronization is therefore critical, for which a common low-frequency reference signal is typically employed across the entire system. Phase shifting devices integrated into the antenna elements can then select appropriate shifts for each transmission device to adjust the beam dimensions for a given task. Beam formation then happens by exploiting constructive and destructive interference between the respective electromagnetic waves. Varying phase shifts and amplitudes also allows the direction of the beam to be controlled, which is typically constrained to a 45-degree window if the antenna surface is planar (Cash, 2021b).

Besides enabling the use of sandwich modules, there are a number of benefits to this power transmission approach. Spatially combining smaller microwave sources offers great scalability in power and distance, which can be tailored for every specific application (Gal-Katziri & Hajimiri, 2018). Additionally, such arrays offer greater beam steering capabilities, added directivity, and enhanced signal-to-interference and signal-to-noise ratios, which also scale with the number of array elements (Stutzman & Thiele, 2012).

So far, phased arrays are already used in radar, sensing, and communication systems (Gal-Katziri & Hajimiri, 2018) but not yet at the kilometer-scale envisioned for SBSP. Typically, they are also bulky, rigid, and heavy (Hashemi et al., 2019). All these qualities would be major disadvantages in SBSP. Hence, there have been first attempts to create large-scale flexible array systems. Experiments using a 16-element 4×4 array powered by integrated geometric concentrator PV cells while operating at around 10 GHz and with an aerial density of 1,000 g/m^2 have been able to successfully power a small LED over a very short distance (Hashemi et al., 2019).

Another potential benefit of light-weight flexible arrays is the possibility of lensing, specifically for planar arrays. Through a combination of the right architectures, circuits, and algorithms, dynamic 3-D lensing of the flexible phased array can result in a higher degree of focus of the microwaves. First tests using free-space dynamic lensing have succeeded in focusing and refocusing of the electromagnetic power beaming field at a distance (A. Hajimiri, Abiri, Bohn, Gal-Katziri, & Manohara, 2020). Using this technique, any field profile permissible by the laws of physics could be created through the correct setting being applied to the individual elements. This could include beam distributions that either form a focal point or maximize the total recovered power at an unknown location in the near or far field. Crucially, the array used for lensing can be arbitrary and non-uniform. This would be very advantageous to the scalability of planar arrays as their shape can vary when elements are added. However, lensing would alter the attitude of the solar side of the array in a non-uniform fashion, the potential effects of which have not yet been investigated.

2.5.2. Optically scanned arrays

So far, the coordination between the antenna array elements has mostly been achieved electronically. The common reference signal is therefore distributed to the modules via cables. These so-called Active Electronically Scanned Arrays (AESAs) are still in active development today for RADAR applications (Yeary, Palmer, Fulton, Salazar, & Sigmarsson, 2021). However, with large arrays such as the ones required for SBSP, electrical synchronization can become challenging. Especially with the shift to more lightweight and thin-film functional elements, the infrastructure for synchronization starts to become one of the main drivers of system cost, mass, and power consumption (Gal-Katziri, Ives, Khakpour, & Hajimiri, 2022). Maintaining timing accuracy for the reference system across a large array electronically also remains a major challenge (Gal-Katziri & Hajimiri, 2018). Hence, electric synchronization mechanisms do not scale well with the array size.

Researchers are therefore looking to replace electronic interfaces with optical connections. The addition of an on-chip photodiode would allow for the timing information to be distributed over long distances via an optical carrier (Gal-Katziri et al., 2022). This optical carrier could either move through fibre or free space. Unlike electric cables, especially the latter option is characterised by great physical flexibility, enhancing modularity and all its benefits. Another benefit would be a reduction of losses within the synchronization infrastructure. For very large arrays, such as a space solar power station, it would also be possible to mix optical synchronization with local electrical synchronization. Overall, optical timing synchronization could therefore be a viable alternative or addition to standard electronic modalities for SBSP by enabling large, scalable, cheap, and lightweight phased array applications. The elimination of cabling could also prove another boost to weight-specific power metrics. Optical information transfer has further been shown to be interoperable with the lensing of flexible arrays (A. Hajimiri et al., 2020).

2.5.3. Phased array algorithms

Lastly, phased arrays need the right algorithm to optimize beam formation and steering. These algorithms need to be adaptive and respond quickly when, for instance, small deformations due to microgravity threaten to exceed the very limited beam steering tolerance from GEO of 0.0005 degrees (ITU, 2021a). Development of such software suitable for SBSP has been progressing recently.

In 2020, the previously 16-element flexible phased array at 10 GHz has been enlarged to 256 elements (Gal-Katziri et al., 2020). By employing a specifically developed algorithm, researchers successfully demonstrated 2-D beam-steering and correction capabilities in response to deformations. The algorithm was also able to focus the beam of a 400-element device and concentrate power on a receiver whose location was previously unknown to the computer. For helical arrays, an additional variable would enter into the algorithm in the form of which of the three dipole antennas of a module should be activated, based on the relative position of the target. Otherwise, the concept remains largely the same.

2.6. Orbits and constellations

With the help of reflectors, a SBSP satellite can achieve 100% duty cycle in GEO by keeping its transmission array earth-oriented. The reflectors then ensure that the solar side of the panels is evenly illuminated, no matter their relative position to the sun. Staying in GEO also ensures that earth does not throw a shadow upon the satellite at any point during the rotation. Such characteristics make it the orbit of choice for all our selected SBSP concepts. However, obtaining GEO requires the satellite components to cover over 35,000 km. Such a great distance results in higher launch costs than closer orbits, which is discussed in more detail in section 4.3. Additionally, as investigated in the following chapter, the greater the distance the larger the microwave transmit and receiver apertures need to be to obtain sufficient efficiencies. Lastly, due to its unique and desirable properties, GEO has already been considered as crowded for a number of years (Jehn, Agapov, & Hernández, 2005). Consequently, researchers have started to look at alternatives that alleviate those problems while maintaining as high of a duty cycle as possible to limit LCOE.

Generally, when a single SBSP satellite moves closer to earth, its duty cycle decreases as the eclipse duration increases and the structure will no longer be stationary relative to the rectenna. As a result, the squint angles required to keep the satellite transmitting for as long as possible increase. Therefore, a single SBSP satellite in a lower orbit than GEO would again be subject to intermittency, potentially requiring grid-scale storage at the ground station. We are therefore facing a trade-off between accessibility, costs, and continuous power supply.

The intermittency issue for closer orbits could be partially overcome by employing a constellation of SBSP satellites. This would ensure that at any point in time, there is a satellite available that is outside the local eclipse and within the maximum permissible squint angle window of the transmit and receiver apertures. MEO would then be the preferred option as configurations would be possible where the individual pass time is still long enough that only few additional satellites would have to be put into operation (Marshall et al., 2021). This is particularly important seeing that multiple space segments would also require more launches. However, MEO also suffers from a much harsher radiation environment than GEO (Larson & Wertz, 1992), increasing the importance of some of the inherently radiation resistant technologies discussed above.

Overall, constellations in MEO consisting of four smaller satellites or more to reduce elevation angles and ground station size while maximising the overall duty cycle can be cost competitive with system in GEO (Marshall et al., 2021). The power-optimal guidance problem from Marshall et al. (2020) could then be extended to N separate space vehicles. Despite multiple satellites also requiring more launch capacity, an issue discussed in section 4.3, they might ultimately be more feasible given the far greater amount of orbits available in MEO.

Furthermore, for planar arrays without external reflectors, the benefit of using dual-sided over single-sided sandwich tiles diminishes as the orbit is decreased and the amount of satellites increased (Marshall et al., 2020). This dynamic is based on the fact that now not a single satellite alone determines the overall duty cycle. Instead a single satellite in position is enough to continue supplying energy to the grid. As a result, satellites could also be smaller while in total providing the same amount of energy. A single satellite at 2 W/g but with a duty cycle of only 50% would, all things equal, provide the same energy as a constellation of smaller satellites with 1 W/g each operating at an overall duty cycle of 100%. Additionally, a constellation could provide power to multiple rectennas on earth simultaneously, given they are spread out enough. This idea is further being pursued as one option for the CASSIOPeiA project to allow for better collaboration

between different governments.

Lastly, as the orbit decreases and the amount of satellites increases, the share of LCOE attributable to the space segment grows (Madonna, 2018). Consequently, the critical metrics for the space segment defined in this chapter become even more important for overall system economics. With the larger squint angles at lower orbits, the required size of the rectenna would also increase. However, this additional cost is insignificant in comparison to the space segment.

2.7. Conclusion

Overall, a number of technology approaches were introduced which are currently not factored into some of the most developed SBSP concepts. Table 1 allows us to directly compare radiation benefits and conversion efficiencies as well as mass-specific power and aerial density values for different parts of the satellite. As explained initially, some numbers from the literature had to be adjusted or extrapolated upon to achieve a like-for-like comparison. In the table, these are shaded in violet. We will begin by summarising the assumptions and the basis upon which they were made. Extensive explanations and references are also provided in the model that was used for the calculations.

To obtain an estimate of the metrics if a new solar technology were to be implemented into a lightweight sandwich module and vice versa, we have calculated a weight premium of 29 g/m^2 which accounts for DC-to-RF and antenna layers. This weight premium is based on an average of the masses reported in Warmann et al. (2020). For dual-sided module designs, a factor of 1.3 was applied to account for shared infrastructure and the mesh design of the optically transparent RF components. An additional weight premium of 32 g/m^2 based on Gdoutos et al. (2020) was used to simulate the effect of supporting and deployment structures for when the sandwich module is integrated into a planar array. Given that we apply the same premia to all technology options, we maintain comparability.

The assumed DC-to-RF efficiency of 70% sits at the lower bound of the investigated concepts and assumptions in the literature (e.g. Sasaki, Tanaka, & Maki, 2013). Furthermore, it is in line with what has been achieved during recent demonstrations for tube-based (Mihara et al., 2018) and semiconductor amplifiers (L.-C. Zhang & Shi, 2022).

When comparing the space segments of the selected concepts, CASSIOPeiA achieves the highest specific power at the lowest overall aerial density. One potential reason for this could be the consequential use of light-weight sandwich modules and high solar concentration levels. In particular, it compares favourably to SPS-ALPHA, whose numbers appear to remain subdued based on heavy module components and external reflector structures. The only concept not employing sandwich modules is MR-SPS, which consequently reports the lowest specific power numbers although not the highest aerial density. The lack of any solar concentration further limits power output. In conclusion, all but CASSIOPeiA fail to beat the baseline comparison CIC and flat-plate cells once transmission infrastructure or more is added. The fact that SPS-ALPHA has a lower p_{PV} than the mass-optimised flat-plate cell despite employing concentrators is likely attributable to the weight of its reflector structures.

Consequently, lightweight concentrator alternatives would be needed. While integrated parabolic mirror modules achieve the highest specific power across all options, in part due to their ultralight design, the lack of commercial availability for the CFRP materials and the complicated production makes their employment at scale uncertain. LSCs could prove to be a viable alternative. However, so far their low PV efficiencies limit specific power results. With convergence of efficiencies, they could become the solar concentration technology of choice due to their low weight, flat-plate geometry, and unconstrained acceptance angles in addition to enhanced radiation resistance.

On the materials side, perovskite has established itself as one of the elements displaying some of the highest p_{PV} values when used in a solar cell. This is despite its PV conversion efficiency only reaching about half that of more established alternatives. Even without any further light concentration, our calculations suggest that it would be able to provide significantly more than 1 W/g, even when integrated into a full array. Its self-healing properties after radiation damage also makes it well suited for the harsh environments of space, particularly when moving from GEO to MEO. Radiation resistance and the resulting weight reduction when other protective measures are removed is also the chief advantage of cells based on nanowire technology. Therefore, nanowires might offer an opportunity to radiation-proof the peripheral components to a perovskite cell and should therefore be used in conjunction with any other radiation-proofing technologies.

Lastly, when comparing our heuristic calculations on dual-sided sandwich modules, it seems that the additional weight would not be too detrimental to their performance. Mass-specific power values fell by about 20%. Therefore, a commensurate boost to their duty cycle would be required to keep mass-specific energy constant. Given that initial studies suggest an increase of 50% to the duty cycle when a second RF layer is introduced, the result would be net positive. Our simulated planar arrays containing these modules also perform at or above the levels of SPS-ALPHA or MR-SPS. However, they are still very early in their development and further tests will have to be undertaken to determine whether such changes to the module are the right approach to solving the power-optimal guidance problem of planar arrays. As a structural alternative, helical designs also come with their own trade-off between a continuous duty cycle and additional weight as well as reduced modularity, flexibility, and scalability. Both helical and planar arrays could then benefit from further mass reductions through at least partially optically synchronised arrays.

In conclusion, the technology alternatives investigated are not exclusively applicable to only one type of satellite. Their utilization in practice should therefore not depend on whether planar or helical structures are ultimately employed. Arguably, all of the concepts might be improved by employing a combination of new technologies. Based on our results, in-

Table 1: A comparison of mass-specific power and aerial density calculations between established satellite concepts and technology alternatives suggests that some innovations could offer notable improvements. Shaded cells indicate that adjustments or extrapolations based on the literature were necessary to obtain a like-for-like comparison. Further indication is given whether a certain technology uses solar concentration or offers radiation resistance. Based on own calculations using the numbers reported in this chapter.

		Solar concentration	Radiation resistance	Conversion efficiency		Specific power [W/g]			Aerial density [g/m^2]		
				PV*-to-DC	DC:RF	PV*	PV* + RF	Satellite	PV*	PV* + RF	Satellite
Concepts	SPS-ALPHA	x		48%	71%	0.95	0.33	0.32	1,598	3,222	3,393
	CASSIOPeiA	x		31%	85%	2.51	1.51	1.51	491	692	692
	MR-SPS†			25%	83%	1.20	0.27	0.16	333	884	1,474
Baseline	CIC			32%	70%	0.54	0.36	0.35	804	833	865
	Mass-optimised flat-plate cell			34%	70%	1.00	0.67	0.62	450	476	508
Integrated parabolic mirrors	7.5 suns low emissivity	x		32%	70%	3.75	2.10	1.72	116	145	177
	15 suns low emissivity	x		32%	70%	5.20	2.59	2.02	83	115	147
	15 suns low reflectivity	x		29%	70%	5.90	2.87	2.15	67	96	128
LSC	Standard	x	x	12%	70%	1.16	0.67	0.57	140	169	201
	Bandwidth optimized	x	x	18%	70%	1.84	1.07	0.90	140	169	201
Cell	Perovskite		x	15%	70%	29.40	2.49	1.26	4	33	65
Nanowire	CIC w/o cover glass		x	32%	70%	1.56	0.99	0.89	278	307	339
	Flat-plate cell w/o cover glass		x	34%	70%	3.00	1.76	1.49	150	179	211
Dual-sided sandwich	CIC with dual-sided RF			32%	70%	0.54	0.28	0.27	804	842	874
	Flat-plate cell with dual-sided RF			34%	70%	1.00	0.51	0.48	450	488	520

* PV includes reflector/concentrator arrays and structures directly attributable to them
† Not based on sandwich module design
Adjusted or extrapolated metric

corporating LSCs into perovskite nanowire cells could result in significant weight reductions, increased inherent radiation resistance enhancing operational lifespan, and notably improved specific power as well as aerial density metrics. These elements could then be incorporated in a sandwich module, given their superior performance when compared to the Chinese variant with separated surfaces. The sandwich modules could then either be used in a planar or helical array and therefore improve the SPS-ALPHA and CASSIOPeiA concept alike. All these optimisation consideration become an even bigger lever for overall system economics if multiple satellites are employed in a lower orbit as their share in LCOE increases markedly.

3. Wireless power transfer and ground structure

Today, we transport energy by various means. Primary energy in the form of oil and gas is hauled across oceans and continents via ships and pipelines. Cables conduct electricity through buildings around the country and under the sea over large distances, connecting grids around the world. WPT follows the same concept of moving energy from point A to point B - just without any physical structures in between. Therefore, a concrete use case might be that point B is a place to which energy is difficult to get to and far away from point A, where energy is abundant, with the area between being ill-suited for wires. This concept also underlies its application to SBSP.

Generally, WPT is defined as "the efficient point-to point transfer of electrical energy across free space by a directive electromagnetic beam" (Rodenbeck et al., 2021) and therefore, in contrast to passive energy harvesting applications, pursues the maximisation of total power transfer efficiency. A standard set-up consists of a transmitting aperture which converts DC power to the electromagnetic waves of choice and which are then sent to a receiving aperture a set distance

away to convert the waves back to DC power.

While the focus of our analysis of WPT remains on its specific applications to SBSP, there is also ongoing research into the idea of energy harvesting, which collects ambient energy present in the form of microwaves to generate electricity (Kazmierski & Beeby, 2014). WPT is also classified as an enabling technology for 6G (Saad, Bennis, & Chen, 2019). Furthermore, we will see in section 3.3.1 that the military has played a critical role in fuelling research behind WPT and continues to do so today, particularly in the United States. While the military always has an inherent interest in space, the strong association of the defence sector with SBSP could pose geopolitical hurdles for its adoption. SBSP satellites could be used to power remote military bases and vehicles (Masrur & Cox, 2019). Theoretically, militaristic applications could also extend to the delivery of disruptive or even destructive electromagnetic power against targets on earth or in space. Therefore, SBSP could be a topic of interest for the UN Committee on the Peaceful Uses of Outer Space to generate operational consensus across nations. These considerations, however, are outside the scope of this thesis.

3.1. Metrics for power beaming evaluation

The following three metrics have been identified from the literature (e.g. Rodenbeck et al., 2022) as key determinants of power beaming systems with a primary focus on microwave power beaming:

Link efficiency [%]:
Link efficiency is often used to describe the share of DC power that is ultimately available at the ground structure after conversion to RF or laser, transmission to earth, collection, and reconversion. Therefore, it is comprised of a collection of efficiencies corresponding to each of these steps. While in theory it would be possible to break the conversion chain down to each individual hardware component, we focus on the sub-efficiencies commonly reported in the literature on WPT demonstrations. This leaves us with three main efficiency metrics for microwaves. First, DC-to-RF efficiency typically includes losses incurred from routing DC power around the space segment and the efficiency of the transmitting antenna. Next, collection efficiency measures the share of emitted electromagnetic waves that ultimately impinges on the reception area of the rectenna. Transmission losses due to interference with the weather and atmosphere are also often included in this metric. At the rectenna, RF-to-DC efficiency is simply the rectification efficiency, including the routing of DC power around the rectenna. For laser beaming, the final step would entail optical-to-DC reconversion via specialised PV arrays. The DC-to-AC conversion required to feed the generated electricity into the grid is not included in any demonstrations but is a standard procedure with relatively high efficiencies.

Frequency [GHz]:
The frequency at which the beam is sent out does not per se determine system performance. However, the chosen frequency has significant implications on overall system design, the scale of the apertures, as well as interactions with the environment. Therefore, it is an important metric when comparing power beaming systems. For optical transmission systems, the wavelength measured in nanometres, which can be easily derived from the frequency, would be the more commonly used metric.

Power density [W/m^2]:
The power density measures how much power is successfully transferred and collected per unit surface area of the receiver aperture. As beams will not be uniform in their power distribution across the surface area, even with proper focussing, a distinction can be made between the average power density across the entire structure and the peak power density incurring only at certain points. The average density is mostly important to the scale of the ground system while the peak density is of particular importance for safety considerations and certification. It can also be expected that power densities will be an important cornerstone in public discussions to build social acceptance, given their close link to system safety.

Regarding the evaluation of demonstrations, it is important to note that inconsistencies in the reported efficiencies of power beaming demonstrations are common (Rodenbeck et al., 2021). Not always are the system boundaries clearly defined and hence some numbers might not be directly comparable. Sometimes certain metrics are omitted completely from the publication. There are some older meta studies that constitute notable exceptions (Brown & Eves, 1992). For future research efforts, it will be crucial that parameter reporting is as uniform as possible with clear definitions of the system boundaries and corresponding hardware components. Additionally, more holistic metrics for evaluation, which may also include costs, have been proposed in the past but not caught on (Dickinson & Maynard, 1999). Nonetheless, efficiencies without such a clear framework, as most of the ones reported in the past, are still helpful to obtain a general understanding of the maturity and progress of microwave power transmission (MPT) or laser power transmission (LPT) technologies.

3.2. Beam types and atmospheric attenuation

The three main types of electromagnetic beams which have been examined for the purpose of power beaming are optical lasers, millimetre waves (mmWaves), and microwaves. Microwaves generally have a frequency ranging from 300 MHz to 300 GHz, whereas mmWaves are a subset of microwaves, representing the higher end of the spectrum at 30 to 300 GHz. On the other hand, laser frequencies are multiple orders of magnitude greater and measured in THz.

For the purpose of SBSP in GEO, these beams have to transport large quantities of energy over distances exceeding 35.000 km. While the majority of their journey will lead them through the vacuum of space, the electromagnetic waves will ultimately also encounter the earth's atmosphere. Here, any beam will experience losses due to interference by the molecules along its path. Coming from GEO, the smallest number of interactions would take place with a receiver

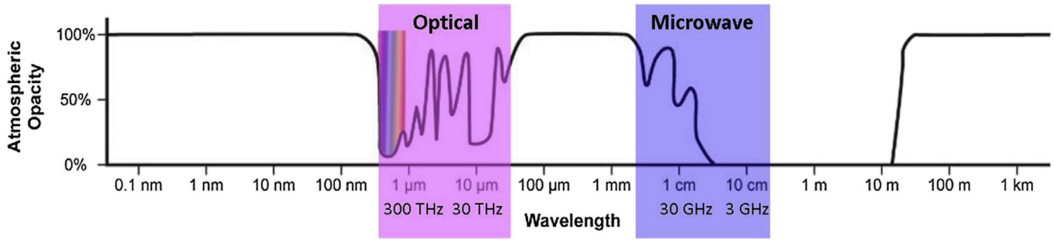


Figure 8: Opacity windows in the atmosphere for optical waves and microwaves. (Jaffe, 2020)

placed directly at the equator, constituting an air mass coefficient of one (AM1). However, the further you move away from the equator, the more atmosphere the beam will have to cross and the more losses it will therefore accrue.

The magnitude of these losses is also closely tied to the wavelength of the beam. Consequently, it is not the entire frequency range of light and microwaves that is relevant to WPT. One of the factors determining the bands of interest are the opacity windows in the atmosphere. These windows represent wavelengths or frequencies where interactions are minimal and are represented in figure 8.

Consequently, to minimise atmospheric losses, frequencies towards the lower end of the microwave band should be deployed, with opportunities for mmWave applications in the K_a- (26.5-40 GHz) or W-band (75-110 GHz). The latter two are highlighted in green in figure 9. These environmental circumstances are further reflected by the fact that the frequencies considered for SBSP are 2.45 GHz and 5.8 GHz (ITU, 2021a). Most demonstrations have employed these levels as well, as we will see in section 3.3.1. In these frequency ranges, atmospheric losses are very small, which makes them well suited for SBSP applications (ITU, 2019).

Converseley, optical applications should use the visible or near infrared parts of the spectrum to utilize the atmosphere's opacity windows. However, energy densities of lasers are by some orders of magnitude higher compared to microwaves (Grandidier et al., 2021), resulting in higher atmospheric losses overall and also poor weather penetration characteristics.

Impediment by weather is another factor of importance when comparing modalities of WPT. While microwave transmission has fewer problems with this as seen in figure 9, lasers tend to be constrained in rainy or foggy conditions. Such issues have also already been observed with other laser applications such as LIDAR (Heinzler, Schindler, Seekircher, Ritter, & Stork, 2019). This presents difficulties for SBSP, given how the underlying concept of renewable baseload power requires the satellite to transmit energy at any time, especially during extreme weather events.

A comparison of all characteristics with respect to the three modalities, including atmospheric penetration, is synthesized in table 2. Overall, one of the chief benefits of optical WPT is the small aperture size. However, despite decent conversion efficiencies, the poor penetration characteristics paired with more stringent safety requirements also resulting from higher power densities present difficulties for SBSP applications. Crucially, optical transmitters at current technological levels are also very difficult to scale due to mechanical tolerances and costs. mmWave applications also suffer from this limitation, despite displaying potentially enhanced safety characteristics, as has been shown in recent demonstrations (Rodenbeck et al., 2021). Nonetheless, this makes these two modalities less suited for the long distances required by SBSP concepts.

In comparison, systems based on microwaves not only perform well on the atmospheric front but also scale in power proportionately to the aperture areas, which constitutes a relatively easy way to enhance system performance. Furthermore, progress in the past decade with vacuum and solid state amplifiers in the antenna allows for higher DC-to-DC conversion efficiencies (Grebennikov, 2011). However, it should be noted that, compared to mmWaves, microwave frequencies correspond to bigger wavelengths, making it very difficult to integrate entire antennas on-chip as has been achieved for mmWave applications (Shaulov, Jameson, & Socher, 2017).

Overall, microwave systems have been favoured for WPT, as evident in the many decades of demonstrations since the middle of the 20th century. Next, we briefly discuss microwave WPT and some of the most significant demonstrations. Nonetheless, we will also briefly show some experiments based on laser transmission and safety-enhancing technologies in section 3.4, as notable progress could be observed in recent years.

3.3. Microwave power beaming

In this section, we start with a brief look at the physical dynamics of WPT systems based on microwaves. Then we will highlight important steps in the history of microwave power beaming, including past and recent demonstrations, followed by spectrum and safety management.

A system to transmit power from a solar satellite in GEO via microwaves would see the generated DC electricity converted to microwaves and send out in form of a beam via an antenna array connected to the orbiting body. As discussed in section 2.5, this transmitting array needs to be phase-calibrated to achieve a concentrated beam for maximum efficiency and targeting precision. The optimal power distribution for a transmit would then roughly resemble the Gaussian curve (Brown & Eves, 1992). The receiving site would convert microwaves to DC using a rectifying antenna (rectenna).

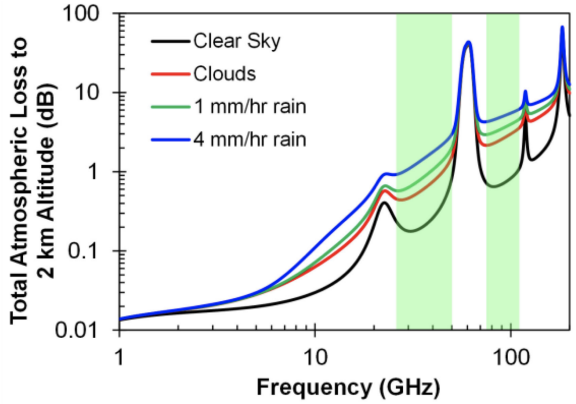


Figure 9: Attenuation effects of different weather scenarios on microwave frequencies. (Rodenbeck et al., 2021)

This rectenna combines an antenna for collection, input filters, rectifying diodes, and an output filter to ultimately feed electricity into the grid (Brown, 1977). Modern versions of such a receiver can take the form of a mesh-like structure. As a result, there are deliberations that a rectenna could be combined with other renewable sources such as terrestrial solar and wind or even employed on agricultural land (Jaffe, 2020).

Integrated rectenna approaches might help with social acceptance, given that the structures would have to be quite large to be efficient. This relationship between size, wavelength, and efficiency for the purpose of SBSP WPT has been expressed by Shinohara (2014), amongst others, via the following formulas. A_t and A_r denominate the area of the transmitter and receiver apertures, respectively, D the distance between the two, and λ the transmission wavelength.

$$\tau^2 = \frac{A_t A_r}{(\lambda D)^2} \tag{12}$$

$$\eta_{Link} = \frac{P_r}{P_t} = 1 - e^{-\tau^2} \tag{13}$$

The metric τ relates these parameters in a way that allows to model link efficiency with the help of equation 13. Both formulas are an adaptation of the Friis transmission equation. Changes are necessary as the Friis variant relies on far-field assumptions that do not apply to WPT (Shinohara, 2014). The far field parameters used for more standard RADAR applications typically do not focus on finite distances. However, maximum energy transfer happens in the radiative near field or Fresnel region, which can still translate to long absolute distances for smaller wavelengths (A. Hajimiri et al., 2020).

The adapted equation demonstrates the trade-off at the core of the system between utility due to size and efficiency. One driver of this dynamic is the collection efficiency, which is the proportion of the radiated waves that ultimately meet the rectenna surface. Any waves that do not reach the receptor result in power being lost. Diffraction physics lead to the simple conclusion that the waves will always be scattered to some degree, requiring very large receiver sizes for the distances encountered with SBSP. Equation 13 does not take these sidelobes into account but could be tweaked by including the beam pattern (Shinohara, 2014).

For an example calculation, a SBSP system in GEO with a 1-km transmit aperture as well as a rectenna measuring 3.5 km in diameter operating at 5.8 GHz could theoretically reach a 90% collection efficiency. It has also been shown that collection efficiencies close to 100% are achievable in practice but require diligent alignment as well as optimised amplitude and phase distributions (Brown, 1974). An even larger transmit array or rectenna could theoretically, within limits, improve this number further. Additionally, should the rectenna not be placed at or near the equator, its required shape would shift to elliptical due to the resulting higher elevation angles of the beam and its surface area would increase further.

On the other hand, the inverse also follows from the equation. As opposed to having large structures in GEO and on the ground to achieve high beam efficiency over long distances, a closer orbit would allow for smaller apertures. However, lower orbits such as MEO would also present two big drawbacks. First, the transmitting and receiving areas would now move relative to each other as the geostationary properties are lost. Hence, some kind of guidance system would be required to maintain the link within the steering tolerance of around 0.0005 degrees (ITU, 2021a). The solution that has been devised to overcome this issue are pilot signals. These are sent from the rectenna to the transmit array and have proven safe and effective (McSpadden & Mankins, 2002). The CASSIOPeiA concept, for example, also plans to employ an encrypted pilot beam to maintain connection for its 1 to 10 GHz WPT system. We will further discuss pilot signals later on. The second drawback is the loss of constant coverage due to relative positioning and effects of earth's shadow. As a result, a single SBSP satellite would no longer be able to provide renewable base-load power continuously. However, a constellation of multiple satellites can overcome this

Table 2: Comparison of the three beaming modalities. Microwaves appear best suited for high-power long-distance applications such as SBSP. Own adaptation based on Rodenbeck et al. (2021)

	Optical	mmWave	Microwave
Weather penetration Clouds/rain/fog	No	Poor	Very good
Conversion efficiency Performance limits for DC/RF conversion	OK	OK	Good
Required aperture size Transmitter/receiver aperture sizes	Small	Medium	Large
Safety Required due regard, pointing, user perception	OK	Good	Good
Economy of scale Present capabilities for high-power/long-distance	Poor	Poor	Good

challenge as discussed in section 2.6.

One factor that has hampered demonstration results from the very beginning until today are the hardware components. Similar limitations such as seen with the Friis equation and underlying far-field assumptions also apply to some of the standard RADAR components used in most WPT experiments. As they are not optimized for high-power beam transmission and collection efficiency at range, achieved link efficiencies can be limited (Shinohara, 2014). One way around this would be customized equipment as used by William Brown during his infamous JPL demonstration, which we will elaborate on shortly.

Overall, there are many trade-offs and challenges when it comes to WPT for SBSP applications. Nonetheless, over the past decades there have been a number of successful demonstrations of the underlying technology using microwaves. In the following section, we briefly introduce the developments and people that led to these achievements, paving the way for today's experiments.

3.3.1. Overview of power beaming demonstrations

In this section we introduce a summary of past, present, and planned microwave power beaming demonstrations and what kind of results they were able to achieve. A generalized representation of a power beaming demonstration, including the key metrics, is provided in figure 10.

Early history

After the concept of transporting electricity in a vacuum without wires was theorised by James Maxwell in 1873 and later theoretically validated by Heinrich Herz around 1890, Nikola Tesla conducted the first experiment with WPT around the turn of the century (Tesla, 1904). Sending alternating surges through masts and thus creating a standing wave between them, he then placed receiving antennas at maximum amplitude points. However, this failed to achieve any significant power transfer (Cheney, 1981). After this unsuccessful demonstration, momentum died down until the middle of the 20th century.

U.S. dominance

With the resurgence of interest in RADAR technology during the second world war, the idea of the wireless transfer of electricity using microwaves also intrigued the military. Further inspired by Isaac Asimov's short story "Reason", which we already introduced in section 1.1 as the first description of SBSP using WPT, the U.S. defense sector started looking into the concept.

At first, focus remained on the possibility of using WPT to power unmanned aircraft for surveillance and communication purposes. In 1959, the defence contractor Raytheon proposed the Raytheon Airborne Microwave Platform (RAMP). The idea was to deploy a small helicopter at 15 km altitude which would act as a communications node and be powered by microwaves from the ground. The required amplitron with an output of 400 kW at 3 GHz and a transmitting efficiency of about 80% was then developed by Raytheon's scientist William Brown in the following year (Skowron, MacMaster, & Brown, 1964). This project marked the beginning of the golden age of WPT demonstrations as well as Brown's position as a pioneer in the field. Shortly after, NASA, who was also involved, successfully improved some key components of the beaming system, allowing for more concentrated beams and hence higher collection efficiencies (Potter, 1961).

In 1963, Brown went on to develop the first complete modern WPT system at Raytheon's lab. The set-up used a magnetron coupled with a reflector to send microwave energy at 3 GHz over a distance of 5.5 m. The resulting DC-to-DC conversion efficiency came in at 16%, based on an 87% collection efficiency and a 50% rectifier efficiency with 100 W of output (Brown, 1980a). Another key element to this experiment was the first modern rectifying antenna to receive the microwaves and convert them back to DC power. Brown was helped by fellow researcher Roscoe George in developing this aperture. Together, they patented the rectenna design a couple of years later (Brown, George, Heenan, & Wonson, 1969). George also conducted his own demonstrations at Purdue based on the design but only managed to achieve 40% RF-to-DC conversion efficiency (George & Okress, 1968). Spurred on by the initial success, addi-

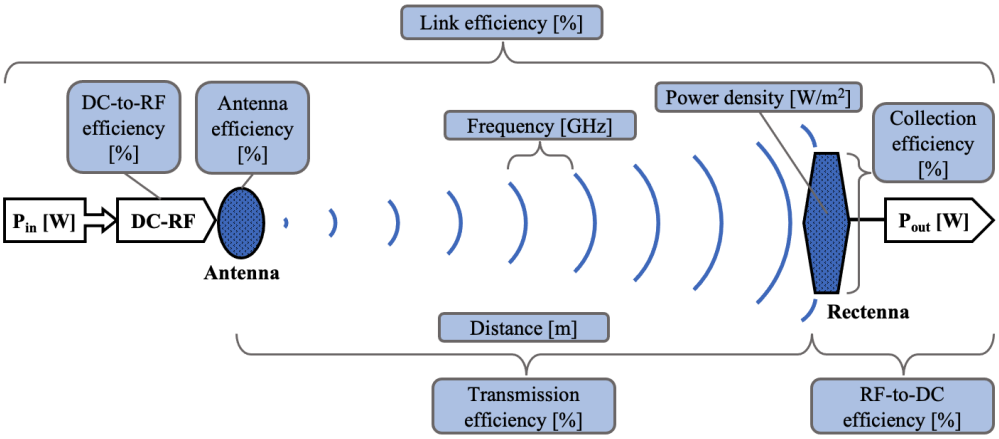


Figure 10: Visualisation of a general power beaming demonstration using microwaves, including the key reported metrics relevant to our analysis. The antenna and rectenna can either be stationary or affixed to moving objects such as terrestrial or aerial vehicles. Own representation

tional contracts were given out to pursue the RAMP concept. Consequently, in October 1964 Brown conducted the first microwave-powered small aircraft flight for which a small helicopter was flown at an altitude of 15.2 m for the duration of 10 hours (Brown, 1965).

Towards the end of the 60s, attention also started shifting towards the possibilities of power beaming in space. Intrigued by Peter Glaser's first concept for SBSP at Arthur D. Little in 1968 (Glaser, 1968), which he further refined in 1973 to elaborate the modalities of the power transfer (Glaser, 1973), NASA gave additional contracts to Brown to push the limits of power beaming at the time. A first demonstration at the Marshall Space Flight Center (MSFC) in 1970 resulted in a measured DC-to-DC efficiency of 26.5% (Brown, 1980a). However, one of the deficiencies uncovered by the experiment was the low rectenna collection efficiency of only 74% versus the theoretical maximum of 100%. As an improvement, the rectenna design was overhauled with elements being spaced more closely and in an overall hexagonal shape. These steps managed to push the initial number up to 93%. Furthermore, the Schottky-diodes essential to the rectification process where switched to be based on different materials. Overall, the rectenna components were also rearranged from its previously flat design to a 3-D volumetric construction to overcome difficulties arising from the new diode spacing and achieve precise polarization alignment (Dickinson, 1975). These changes also boosted performance to the point that another demonstration at the MSFC in 1974 yielded a far improved RF-to-DC efficiency of 82%, resulting in an impressive DC-to-DC efficiency of 48% (Brown, 1980a).

Activities then shifted from the MSFC to the Jet Propulsion Laboratory (JPL) and the NASA Goldstone Deep Space Communications Complex (Goldstone) in California. At this point, Richard Dickinson from the JPL also became very involved in the experiments. Ultimately, the efforts at the JPL culminated in a demonstration in 1975 where power was beamed over a distance of a couple of meters via microwaves at 2.45 GHz, delivering DC power of 495 W at the rectenna (Dickinson & Brown, 1975). Conversion efficiencies of 69% and 79% for DC-to-RF and RF-to-DC respectively, resulted in an overall strong system efficiency of 54%. This number constitutes a benchmark that stands unbeaten to this day, despite Brown identifying technical potential to boost link efficiency further to 76% in the wake of the experiment (Brown & Eves, 1992).

Later the same year, another demonstration took place at Goldstone. Based on the improved volumetric design, the largest rectenna array to date was built, spanning 24 m^2 with performance still exceeding 80% efficiency (Dickinson & Brown, 1975). It was used to receive microwave energy sent from an antenna 1.54 km away at a frequency of 2.4 GHz, resulting in 30.4 kW DC power output with incident peak RF intensities of up to 170 mW/cm^2 (Dickinson, 1975). It remains the highest power result to date, although the end-to-end efficiency stood at only 4% as it was limited by the transmission and receiving apertures (Dickinson, 2003).

In 1976, Brown followed up his success by tweaking the rectenna design using a custom air-metal diode technology which unfortunately has since been lost (Rodenbeck et al., 2021). While its power handling capabilities were only in the 10s of watts, it helped him achieve the highest reported RF-to-DC efficiency to date of 91.4% (Brown, 1977). This result serves as an example of what could be possible with hardware customised for WPT. Nonetheless, due to the difficulties in fabrication, the 3-D volumetric rectenna design was abandoned in the 80s in favor of a thin-film variant using photolithographic techniques. The thin-film design was also able to demonstrate RF-to-DC conversion efficiencies of over 80% at 2.45 GHz (Brown & Triner, 1982).

The 1970s thus turned out to be the golden age of WPT demonstrations. The rapid progress increased NASA's confidence in the technology and its applications for SBSP. Further theoretical studies in collaboration with the U.S. Department of Energy followed, which culminated in a report concluding

that SBSP with WPT was a feasible technology for the future (Brown, 1980b). The 670-page document was also the first time the concept of retrodirectivity to keep the beam on target was introduced. However, for reasons unknown, the publication marked the conclusion of NASA-sponsored programs and hence the end of U.S. leadership in the field of WPT.

While the U.S. transitioned into a passive role, countries like Canada continued with their own research programs. The Stationary High-Altitude Relay Program (SHARP) culminated in the 20 minute long flight of a lightweight fuel-less airplane (Schlesak, Alden, & Ohno, 1985). Achieving altitudes of up to 150 m, it was remotely powered by microwaves at 2.45 GHz and power densities of up to 400 W/m^2 were measured at the wing. SHARP was also one of the first experiments to make use of the light-weight thin-film rectenna devised by Brown just a couple years earlier. Their properties made them particularly well suited for airborne vehicle applications.

Japan's advances

In the end, it was Japan who really took over the mantle of leadership in WPT research. As a country of few natural energy resources yet with a lot of high-tech industry, it was quickly enamoured by SBSP and its promises of some clean energy self-sufficiency. Hence, the Japanese government, universities, and businesses became quite active in the pursuit of key technologies needed for SBSP - particularly WPT. It is notable that, unlike in the U.S., the private companies came from outside the defence sector. In its pursuit, Japan continued to push the boundaries and was the first to conduct in-space experimentation in 1983. The Microwave Ionosphere Nonlinear INteraction eXperiment (MINIX) was designed to gain insights into the interactions between ionospheric plasma and high-powered microwaves (Matsumoto et al., 1982).

Another valuable contribution by Japanese researchers was the technical development of beam tracking. With the idea of retrodirectivity introduced in the 1980 U.S. report, it was scientists at Kyoto University in 1987 who collaborated with Mitsubishi Electric Corporation to develop a retrodirective transmitter for automatic beam alignment between transmitter and receiver (Matsumoto, 1989). The concept was based on the transmitting array sending the electromagnetic waves in the direction of a pilot signal operating at a different frequency. Such targeting would not only be crucial for systems operating outside of GEO, but small gravitational deformations in an array, such as introduced in section 2, could also have adverse effects on beam accuracy. The technology was then further improved upon with Nissan Motor Company in the mid 1990s.

The automatic alignment technology opened up new possibilities for experimental set-ups. In 1992 during the Microwave Lifted Airplane eXperiment (MILAX), Japan ran a successful demonstration of microwave power beaming between two moving apertures (Fujino et al., 1993). An electronically scanned phased array mounted on a moving vehicle was successfully used to focus a 2.4 GHz beam on an airplane in motion relative to the ground-based aperture.

Emboldened by its quick scientific advances, Japan was then the first country to achieve power beaming in space. Its International Space Year - pulsed Microwave Energy Transmission in Space (ISY-METS) experiment on 18 February 1993 saw microwave energy beamed between two rockets during launch (Kaya, Kojima, Matsumoto, Hinada, & Akiba, 1994). One rocket was carrying microstrip antenna arrays and the other different rectennas, one of them designed in the U.S., between which microwaves were then pulsed (Akiba, Miura, Hinada, Matsumoto, & Kaya, 1993).

Additional experiments were then undertaken in the mid 1990s. The first was to better understand the dynamics of large rectenna arrays consisting of many elements and the required characteristics of each element as well as how to connect them to maximise DC output. For that reason, researchers from Kyoto University worked together with Kansai Electric Corporation to target a beam across a distance of 42 m at different rectenna constellations (Shinohara & Matsumoto, 1998). Optimal results were achieved when rectennas of equal DC output were connected. However, for the huge arrays necessary for SBSP it will be impossible to have equal microwave power densities across the entire aperture surface and very difficult to mass-manufacture elements with identical performance. We will further discuss manufacturing briefly in section 4.3. The second was another airship experiment called Energy Transmission toward High-altitude long endurance airship ExpeRiment (ETHER). A blimp equipped with a 9 m^2 rectenna array consisting of 1,200 elements was flown at an altitude of 50 m for four minutes (Kaya, 1996). The aircraft was powered by a microwave beam delivering 10 kW of power at 2.45 GHz. In line with most past experiments, RF-to-DC conversion efficiencies of 81% were achieved.

During the course of its MILAX and ETHER experiments, Japan also introduced dual polarization to its rectenna designs (Fujino et al., 1993). This allowed the transmitting and receiving array to be rotated relative to each other while maintaining stable DC power output. Hence, the two apertures no longer absolutely needed to be completely parallel to each other. This property would be crucial for any SBSP system design that either had the rectenna outside the equator area or any satellite not in GEO. The latter is the case for Japanese SBSP concepts as shown in section 1.2.

Despite the lack of engagement by key institutions such as NASA, some U.S. universities still conducted their own demonstrations during this period. For example, the University of Alaska Fairbanks constructed a Semi-Autonomous BEam Rider (SABER) helicopter in 1995 (Hawkins, Houston, Hatfield, & Brown, 1998). It was powered by a 1 kW transmitter operating at 2.45 GHz and showcased at the WPT conference in Japan.

International momentum

Over the next decades, momentum for MPT increased around the globe. Encouraged by Japan's advances, NASA

performed another comprehensive evaluation of the prospects of SBSP, the positive results of which prompted it to initiate the SSP Scientific Exploratory Research and Technology (SERT) program in 1999. The goal was research on key space solar technologies, including power beaming. As part of this program, a public demonstration was conducted in 2002 at the World Space Congress in Houston by targeting a 5.8 GHz beam at a 1 m rectenna array (Strassner & Chang, 2003a). The system was also equipped with a retrodirective transmitter, allowing the beam to track the receiving aperture while it was moving around the exhibition floor. A RF-to-DC conversion efficiency of 82% was ultimately achieved, once again in line with past results but still lagging behind Brown's record from 1976. At the same congress, Japan exhibited the first fully integrated solar power radio transmitter (Matsumoto, 2002).

Around the turn of the millennium, South Korea also launched its own WPT program through the Korea Electrotechnology Research Institute (KERI), which, based on private communications with external scientists, managed to demonstrate a total power beaming efficiency of 44% with a final DC power output of just over 1 kW (Rodenbeck et al., 2021). If correct, this would be the first result that at least got close again to Brown's 1975 JPL experiment.

While the World Space Congress helped restore momentum to the WPT ambitions of the west, Texas A&M University continued innovating the rectenna with a printed circular polarization design (Strassner & Chang, 2003b). The new approach achieved 78% RF-to-DC efficiency at 5.6 GHz and allowed for 4 times fewer diodes over the surface of the rectenna compared to old designs. Such improvements directly translate into simpler manufacture and overall lower costs, both of which crucial factors for SBSP. At the same time, other American universities tested wideband rectenna arrays to potentially achieve flexibility regarding the transmission frequency but could not achieve significant conversion efficiencies (Hagerty & Popovic, 2001).

Also during this period, while JAXA published its SBSP concepts, Kyoto University developed its Space POwer Radio Transmission Systems SPORTS-2.45 and SPORTS-5.8 (Shinohara, Matsumoto, & Hashimoto, 2004). By replacing the standard microwave tubes in the transmitter with phase-controlled magnetrons (PCM), the researchers achieved higher efficiencies at higher kW power levels while also improving beam steering capabilities. This PCM technology was then used in 2008 for a relatively impromptu yet historic joint Japanese-American demonstration in Hawaii (Foust, 2008). Using microwaves, researchers successfully beamed power across the 148 km distance separating Maui from the big island of Hawaii. While beam collection efficiency remained very low at less than a thousandth of a percentage point and power had to be kept low as not to interfere with air traffic, the significance of the demonstration was showing that power could be sent via microwaves across a distance roughly resembling the depth of the earth's atmosphere. Despite the PCMs, keeping the beam on-target proved one of the key challenges. One year later, Kyoto University further tested two 110-W PCMs by beaming power at 2.46 GHz from a blimp to the ground (Shinohara, 2013). The system was also using a pilot signal for beam tracking and reached transmitting antenna efficiencies of 54.6%. It was also around this time that the idea of transparent antenna arrays as discussed in section 2 started taking hold.

Microwave power beaming today

As momentum continued to build in the last decade, Japan succeeded in multiple large power beaming field experiments as part of its SBSP initiative by the Ministry of Economy, Trade and Industry (METI). Once again displaying the deep involvement of the Japanese private sector, in 2015 Mitsubishi Heavy Industries transmitted 10 kW at 2.45 GHz with an aperture separation of 500 m (Nishioka & Yano, 2015). Thus, they set a new Japanese record for distance and power transmitted. The corresponding efficiency numbers have not been reported. Also in 2015, METI conducted a horizontal power beaming demonstration at the Mitsubishi Electric Facility (Mihara et al., 2015). 1.8 kW were beamed 55 m at 5.8 GHz with a link efficiency of 18.6%. This particular experiment was preceded by lab testing undertaken by J-Space System (Takahashi et al., 2016). The researchers successfully showed that accurate beam targeting could be maintained with the help of a retrodirective subsystem despite temporary misalignments of the antenna panels. As discussed before, such gravitational misalignments would have to be expected for any array-based SBSP system. The most recent Japanese demonstration was held in 2019, where microwave power was successfully beamed at a density of 4 kW/m^2 towards a drone at a distance of 10 m (Shinohara, Hasegawa, Kojima, & Takabayashi, 2019). The 60 W of power delivered, reduced to 42 W at 30 m altitude, successfully prolonged the battery life of the drone.

Today, with a focus on sandwich modules as discussed in section 2, Japan is looking to develop better Schottky diodes to improve efficiencies of the rectification process (Mizojiri et al., 2019). This can be seen as a promising approach, given what Brown was able to achieve with his custom-built diodes. Japan also initially had its first small solar satellite WPT demonstration planned for 2015 (Tanaka, 2021). While it has been delayed due to a lack in technological maturity of WPT systems, amongst other things, Japan still has ambitious goals. A key to their strategy remains the continued development of daily-life use cases together with industry partners for key SBSP technologies such as WPT. JAXA has also communicated plans for more long-distance demonstrations, which will then inform a decision in 2025 on whether to proceed with a demonstration involving a full system in space.

In the meantime, China has also launched multiple WPT research programs at different universities across the country. With its typical ambitious and centrally planned approach, it expects to become the first nation to build a space solar power station with practical value as announced through the China Academy of Space Technology (Lei, 2019). In pursuit of this goal, China has been building a dedicated Space Solar Experiment Base in Chongqing covering 130,000 m^2, which

will be open to international experts.

Chinese programs have already produced a number of successful demonstrations. In 2014, researchers at Sichuan University beamed a small amount of power via microwaves at 2.45 GHz across 4.5 m, achieving a link efficiency 14.2% (Rodenbeck et al., 2021). Two years later, they experimented with a subarray decomposition of the rectenna at 5.8 GHz, boosting overall efficiency by more than 10% (H. Zhang & Liu, 2016). In 2018, a demonstration in Xi'an also achieved 66.5% RF-to-DC conversion efficiency while using the first focused MPT system with circular polarisation (Dong et al., 2018). During the same year, plans were announced to build an entire SBSP system on the ground for testing, with a targeted beam distance of 100 m, power output of 1 kW, and DC-to-DC efficiency of 20% (Hou & Li, 2021). This project has not been concluded yet. The following year, similar targets were achieved by researchers at Wuhan University at the less commonly used frequency of 10 GHz (Rodenbeck et al., 2021). An approximate link efficiency of 19.5% across 100 m was achieved. The most recent known Chinese demonstration took place at Sichuan University in 2020, with a system operating at 5.8 GHz at 18.5% DC-to-DC efficiency over a distance of 10 m (Chen, 2020). China has also contributed to rectenna designs by improving their heat management, enabling pferomance at higher power levels (B. Zhang et al., 2015). Overall, China has been able to achieve quite impressive results comparable to other nations, despite having started its efforts rather recently. Nonetheless, the Chinese MR-SPS concept targets a WPT efficiency of 54%, in line with Brown's record. Therefore, demonstrated efficiencies still need to nearly triple to reach that goal.

For the future, China plans to launch tethered balloons with solar panels at its newly built experiment base (Lei, 2019). Once those have reached their operational altitude of 1 km, they will collect energy from the sunlight and beam it down to the ground. There have also been plans for another low-power demonstration at 5.8 GHz across a distance of up to 100 m at the same base (Rodenbeck et al., 2021). However, no updates are available as of yet.

The country of South Korea has also continued its engagement with MPT. A notable advancement are Korean efforts towards better heat management and high-power performance of rectennas in 2018 (Park, Kim, & Youn, 2018), a continuation of previous research by the Chinese. Additionally, in 2019 the U.S. entered into a partnership which resulted in another drone demonstration (Song et al., 2019). 10 GHz of power was beamed at an airship at a NASA facility using 32 rectenna array sheets in total and achieving speeds of 7 mph. In line with this experiment, Korean companies are working on microwave power transmitters in alternative ranges such as the X-band at 7 to 11 GHz (Rodenbeck et al., 2021).

Meanwhile, the U.S. published its D3 Space Solar Proposal in 2016, declaring its intention to become the leader in SBSP (SDSC, 2016) - possibly also in response to Chinese advances. Agencies involved include the Department of State, Department of Defense, DARPA, U.S. AID, NRL, the Air Force, and defence contractor Northrop Grumman. As mentioned in section 2.3.1, the United States has also conducted its own space tests with the PRAM-FX, during which DC-to-RF efficiencies of 37.1% could be demonstrated (Rodenbeck et al., 2021). These results outperformed previous ground tests of the employed sandwich modules (Jaffe, 2013). The California Institute of Technology is another American organisation deeply involved in research on modular phased arrays, as noted in section 2, including timing devices to achieve synchronization at large scales for focused microwave beams (Gal-Katziri & Hajimiri, 2018).

Recently, another MPT demonstration has taken place in the U.S. to test the practicality of terrestrial microwave power beaming over distances exceeding 1 km (Rodenbeck et al., 2022). As this specific set-up made use of ground bounce properties of microwaves over cluttered terrain to boost power density and efficiency, it is not fully applicable to SBSP use cases. Due to limited aperture sizes and power handling capabilities of the utilized commercially available diodes, which required beam defocusing, overall efficiency was initially limited to 5%. Nonetheless, the ensuing case studies have shown that such a setting could achieve link efficiencies of up to 44%, e.g. by increasing rectenna aperture areas by a factor of 20. This is a relevant result for SBSP, as aperture areas are required and expected to be very large.

Over the next couple of years, the Air Force Research Laboratory (AFRL) will be managing the Space Solar Power Incremental Demonstrations and Research (SSPIDR) project (Rodenbeck et al., 2021). The approach consists of incremental demonstrations and further development of key technologies, including MPT, and is structured in four phases. Ideally, these would then culminate in a fully operational SBSP constellation. The current phase one plans for three major demonstrations. The first, called Arachne, attempts to be the world's first space-to-ground power beaming demonstration by a modular sandwich panel with integrated beam formation optimisation via in-situ array shape measurement and is slated for 2023. The second, SPIRRAL, will test thermal management capabilities of the system and is also planned for 2023. Finally, SPINDLE will test the overall orbital structure deployment. These ambitious projects reflect the opinion of leading American researchers on MPT that the technology has progressed enough to allow for real world developments, as expressed by Rodenbeck et al. (2021).

There are also private companies active in the field of WPT without any direct governmental involvement. For example, Emrod from New Zealand is planning to commercialise MPT (The Economist, 2021a). In collaboration with Powerco, a local electricity distributor, a prototype WPT system has been developed in an enclosed test facility. Next, Emrod plans to beam power in the kW-range from a solar farm to a client some 2 km away. The company claims an efficiency of about 60% and intends to boost this number further by using relays to refocus the beam along the way. This, however, will not be an option for SBSP. Another private sector institution involved is Solaren. The California based company aims to develop commercial SBSP plants to ultimately operate them and sell their electricity (Solaren, 2022). For this purpose,

they are also researching WPT systems but have not publicised any demonstrations so far.

In conclusion, there have been a number of successful MPT demonstrations over the years, addressing some of the difficulties a system for SBSP at scale might encounter. Great interest in the possibilities of this transmission approach has been evident across countries and driven by a diverse set of actors. However, technical implementation is not the only challenge for power beaming based on microwaves.

3.3.2. Spectrum management

One of the bigger challenges for WPT remains spectrum management. Globally, the International Telecommunications Union (ITU) allocates frequencies for different uses (ITU, 2021b). So far, no wavelength in the spectrum has been assigned to microwave power beaming. It would require a long and arduous process through the ITU to achieve that. Additionally, national bodies such as the Bundesnetzagentur (BNetzA) in Germany or the Federal Communications Commission (FCC) and National Telecommunications and Information Administration (NTIA) in the U.S. impose their own rules for spectrum utilisation. This results in a difficult environment for new technologies to gain a foothold and can hamper development and innovation.

Nonetheless, patterns regarding favoured frequencies have emerged over the past decades. As we have seen, most demonstrations take place around the 2.45 GHz and 5.8 GHz frequencies. Both of them reside within the Industrial, Scientific, and Medical (ISM) frequency bands as classified by the ITU (ITU, 2021b). However, this broad definition is resulting in the bands getting crowded. Hence, there was a notable shift in some recent experiments towards higher frequencies around 10 GHz, also known as the X-band. As discussed at the beginning of this section, the atmospheric properties of the X-band are slightly worse yet still comparable to those of lower frequencies.

With this range of possible options, standardisation is key. So far, no regulatory definition of power beaming services exists to clear the path for more focused technological development in harmony with regulatory ambitions. Only Japan is currently making strides to tackle the lack of regulatory support by trying to establish WTP standards through the Wireless Power Transfer consortium for practical applications (WiPoT) and the Broadban Wireless Forum (BWF) (ITU, 2021a).

Potential interference with other services also remains a prominent issue. For example, telecommunication companies in the U.S. do not require licenses to operate in the ISM band. However, they are not allowed to interfere with other devices in the same band, which can limit the effectiveness of power beaming in the ISM. This was the case during the 2008 Hawaii demonstration, where power levels had to be kept subdued, resulting in lower efficiencies. Besides the base frequencies used for WPT, harmonies up to the 10^{th} level also need to be filtered to prevent interference with devices operating at these frequencies. This includes other satellites as well as some of the most restrictive bands reserved for radio astronomy (DoC & NTIA, 2021).

Rodenbeck et al. (2021) have conducted a simulation using the 2.45 GHz (NASA, 1978) and 5.8 GHz SBSP reference systems by NASA (Davis, 2012). As shown in figure 11, sidelobes with significant power levels remain even at high collection efficiencies. As a result, many hundred MW of power would be scattered away from the rectenna, degrading sensitivities of Bluetooth devices or radios for thousands of kilometres away from the rectenna. Harmonics of 2.45 GHz and 5.8 GHz also fall within primary space-to-earth service bands (ITU, 2021b). Consequently, transmitting antennas would have to filter those harmonics down to non-interfering power levels, as existing communication systems are not adapted to deal with these levels of potential interference.

3.3.3. Safety

Based on past microwave technologies, safety thresholds for microwave exposure have already been defined by various organisations. In general, the Institute of Electrical and Electronics Engineers (IEEE) considers a power density of about 100 W/m^2 safe for controlled access areas (IEEE Standards Coordinating Committee, 2019). This number does not necessarily constitute a hard physical limit, given that power transmitted by the sun's light can reach 1000 W/m^2 in certain places during the summer (Koblin, Krüger, & Schuh, 1984). However, this increased limit would still be lower than some of the power densities achieved in the past. For example, the 1975 Goldstone demonstration peaked at 1,700 W/m^2 (Dickinson, 1975).

Hence, systems must be in place to ensure that adequate limits are not exceeded and that the beam remains on target. The necessary steps are often referred to as the "6 Ds" (Rodenbeck et al., 2021). In particular, they specify that a system should be able to Detect any potentially unsafe situation in order to then Decide whether to Defocus, Divert, Dim, or Douse the beam. Nonetheless, most demonstrations have stayed well within these limits and diffraction physics also mean that over the large distances SBSP requires, power will inevitably spread out. The latter would be conducive to lower power densities as measured in W/m^2, suggesting the clear possibility of a safe MPT environment. Once again, uniform standards appear key to develop the technology towards a safe direction.

3.4. Laser power beaming

Microwaves are without question the modality of choice for all major power beaming concepts. Nonetheless, laser power beaming made some recent progress and is hence included in this analysis as a somewhat possible alternative. A chief benefit of LPT is without a doubt the smaller scale of the systems involved. Given the much smaller wavelength when compared to microwaves, the sizes of transmit and receiver apertures are a fraction of those of MPT systems. For instance, an optical beam near the infrared spectrum at 795 nm would require a receiver of only a few meters in diameter

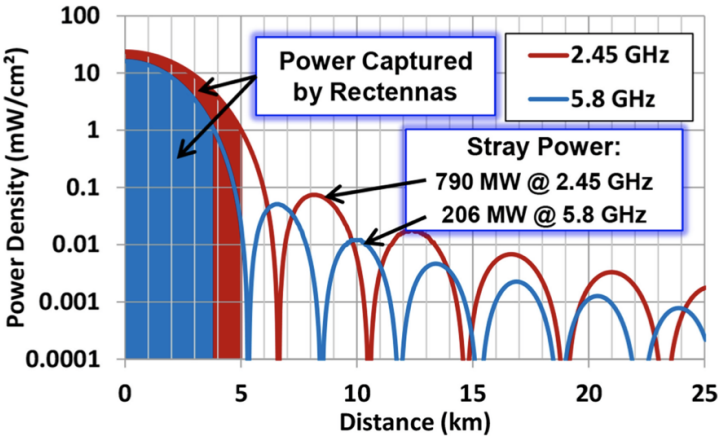


Figure 11: Sidelobe patterns for the 2.45 and 5.8 GHz NASA SBSP reference systems. (Rodenbeck et al., 2021)

(Rubenchik, Parker, Beach, & Yamamoto, 2009) as opposed to the kilometer-scale rectennas used for microwaves. These considerations were first introduced into the discussion in the early 2000s (Penn, Law, et al., 2001).

Additionally, laser beaming from space could make use of so-called direct solar pumped lasers, which use concentrated solar irradiation directly as an energy source for the beam formation in the gain medium without the need of going through an electrical conversion first (Summerer & Purcell, 2009). This bears the potential of lower conversion losses and the avoidance of high voltages on the space segment. However, they prove difficult to scale to SBSP levels and heat management of the laser generating segment remains an issue.

On the receiver end, PV cells specifically designed for the transmission wavelength of the laser could be used. Therefore, these cells can be fine-tuned for a small portion of the spectrum instead of for the majority of it like standard PV cells (York & Fafard, 2017). Such specialized multijunction cells are pushing monochromatic optical-to-DC conversion efficiencies close to 70 %, approaching those of rectennas.

Due to the long distances across which power must be transferred for SBSP, fibre lasers promise to be a well-suited variant based on their high levels of brightness and beam quality (Grandidier et al., 2021). Overall, they are characterised by superior heat management when compared to conventional lasers, high efficiencies due to the use of laser diodes, and good scalability of optical power. As a result, conventional fibre devices can achieve link efficiencies in the range of 20% to 30%.

Finally, beam tracking is also of great importance to LPT systems, particularly because the target is much smaller. The same retrodirective approaches introduced for microwave beams and involving a pilot signal could also be applied to lasers (Summerer & Purcell, 2009). The concept of array building to overcome hardware scalability issues is likewise a theoretically viable alternative for lasers.

3.4.1. Selected demonstrations

So far, there have been a number of notable laser power transmission demonstrations, particularly in recent years. In 2004, Japanese scientists developed a set-up intended to use a laser to beam power to a rover looking for ice on the moon (Kawashima & Takeda, 2004). In a terrestrial demonstration with a life-size rover, power was beamed across a distance of 1.2 km with a link efficiency of more than 20%.

Laser transmission was also investigated in the context of powering robots or aerial vehicles. One such demonstration using a 200-W laser at 808 nm achieved output power levels of 40 W at the target, resulting in a link efficiency of 20% (Kawashima & Takeda, 2008). However, the large amounts of dissipated heat when output power was increased above a couple hundred watts and the lack of compact cooling systems suitable for such high-power lasers remained an issue.

More recently in 2014, a high-power transmission system was demonstrated by Chinese researchers for a distance of 100 m and employing a 793-nm wavelength (Tao et al., 2014). Using multijunction GaAs cells for reconversion, the optical-to-DC conversion efficiency remained subdued at 40 %, resulting in a link efficiency of 11.6%. Nonetheless, a power density of 60,000 W/m^2 was measured. Such high levels are indicative of why safety is of an even higher importance for laser-based transmission systems. We briefly explain some safety thresholds and mechanisms in the following section.

Japan has also conducted a laser power transmission demonstration in 2016, in which a beam was sent from the top of a tower to a receiving station on the ground, thereby bridging a vertical distance of 200 m (Tanaka, 2021). At 1070 nm and an output power of 350 W, 74.7 W were ultimately received on the ground. This results in a link efficiency of about 21%.

The most recent and most advanced demonstration took place in 2019 at the U.S. Naval Surface Warfare Center in Maryland (NRL, 2019). In collaboration with the private company PowerLight Technologies, the NRL conducted the

Power TRansmitted Over Laser (PTROL) project. It consisted of a 2-kW laser transmitter and a receiver specifically optimised for the laser's wavelength. Ultimately, 400 W were beamed across a distance of 325 m and converted back to AC to power lights, several laptops, and a coffeemaker. A key challenge during the project included interference with the weather, such as snow or rain. Unfortunately, the achieved link efficiencies were not openly communicated.

3.4.2. Safety

Similar to microwaves, there are organisations which have defined exposure limits for optical beams. For instance, the Laser Institute of America (LIA) publishes a series of safety guidelines and standards through the American National Standards Institute (ANSI). Recent standards put the limit at 1,000 W/m^2 (The Laser Institute, 2014). However, as was demonstrated during the Chinese experiments, these limits can be exceeded. Especially in the military, the threshold is commonly crossed if other safeguards are in place that make it highly unlikely for anyone or anything to be exposed to the beam directly.

One innovative approach for such a safeguard system was also on display during the 2019 demonstration by the NRL. By employing guarding sensors which figuratively caged the beam, objects could be detected before they reached it and the system would be turned off (Nugent, Bashford, Bashford, Sayles, & Hay, 2020). It could then restart automatically in a few seconds after confirming that there were no longer any foreign objects in the beam's path.

3.5. Conclusion

Across the different power beaming modalities, microwaves seem to have been established as the technology of choice. While lasers offer the possibility of smaller aperture sizes and hence decreased space segment mass, their poor weather penetration characteristics and lack of cost-efficient scaling pose significant hurdles to their implementation in SBSP systems. Consequently, all major advanced concepts so far rely on MPT.

For microwave-based approaches, there have been many successful demonstrations over the past decades. These prove the technical practicality of the technology. Still, none of the distances so far have come close to what would be required of a space-to-earth WPT. Consequently, the scale of transmitting and receiving apertures has also remained limited compared to the km-size structures necessary for SBSP. A comparison of orbital distances for different concepts and the maximum achieved can be found in table 3.

Many of the records for beaming efficiencies established by Brown in the 70s and 80s still stand to this day. Specialized hardware tailored to MPT instead of general RADAR applications has played an important role in his achievements and will be required again to maintain the efficiency levels necessary for SBSP. However, the overview in table 3 also shows that the conversion efficiencies achieved during demonstrations are already in line with what major SBSP concepts expect. The persistent misconceptions about its feasibility therefore seem unfounded and based on misconceptions regarding the technology and associated far-field assumptions (Shinohara, 2014).

The contribution of the WPT segment towards LCOE also mostly comes from the amount of energy that is transmitted and the cost of the rectenna. Variance analyses for LCOE so far suggest that the latter does not have a significant impact relative to the other cost positions (e.g. Way & Lamyman, 2021a). Therefore, a trade-off between an increased aperture size to maximise microwave collection efficiency and increased construction costs might be beneficial.

Overall, as space is getting more commercial and political, momentum for WPT in a space setting is increasing. Another indication for this dynamic is the most recent wave of in-space demonstrations and China as a notable new entrant into the field. Especially the U.S. will see itself pressured not to be left behind by ambitious Chinese WPT targets. Nonetheless, it is critical for the main actors to define a common regulatory definition of the technology around which standards could be formulated. Such shared frameworks would help overcome the remaining challenges regarding spectrum and safety management and include starting or supporting the necessary processes within the ITU. Furthermore, early engagement with society at large to educate them on WPT could help drive the social acceptance of the technology.

4. Infrastructure and manufacturing

Having covered the space segment as well as the energy transfer to earth, we now take a look at the production and infrastructure necessary for SBSP. During our brief analysis of the manufacturing side, a particular focus will be placed on economies of scale and first considerations regarding the life cycle of the system. For the required infrastructure, ongoing developments in the launch market play a critical role in determining the technical and economic feasibility of SBSP. As part of this, we will also discuss some of the orbital considerations and implications of the chosen orbit on capacity and cost. Overall, launch infrastructure and capacity will be the focus of this chapter.

4.1. Metrics for infrastructure and manufacturing evaluation

For the production, launch, and orbit of SBSP systems, the following metrics were determined to be of importance to gauging overall system performance.

Mass-specific costs [$/kg]:
This metric measures the costs incurred per unit mass of the SBSP system to render it operational and can relate to a number of processes. First, the average production costs of the sandwich modules and other components for the satellite as well as the rectenna can be measured in a comparable way

Table 3: Comparison between the MPT metrics targeted by major SBSP concepts and what has been achieved during power beaming demonstrations. While no distances in the same order of magnitude have been achieved, transmission efficiencies are already at or above the desired levels.

	Beaming modality	Frequency [GHz]	Efficiencies				Distance [km]	Power density [W/m2]
			DC-to-RF	Collection*	RF-to-DC	Link		
Concept target metrics								
SPS-ALPHA	Microwave	2.45	70.6%	83.3%		58.8%	36,000	85
CASSIOPeiA	Microwave	2.45	85.0%	82.0%	85.0%	59.2%	36,000	60
MR-SPS	Microwave	5.80	83.3%	90.0%	76.5%	57.4%	36,000	73
Demonstrated values								
Maximum achieved	Microwave	2.45 - 10.00†	>70.0%	>90.0%	91.4%	54.2%	148	1,700

* Based on diffraction physics, mostly depends on relative antenna/rectenna dimensions, distance, and beam steering capabilities

† Range of all performed demonstrations

using mass-specific costs. Second and arguably more importantly, launch costs are measured in dollars per kilogram payload. Overall, this metric can then be used to put a dollar value on some of the calculations from section 2.

Launch capacity [t/a]:
The launch capacity per year denotes the total mass that is transferred into space throughout a certain time period, for instance a year. Due to the long lead times in the launch market, this number can either be based on past realised launches or services already locked in for the near future. Given the great mass of all SBSP satellite concepts, launch capacity is critical in determining the time scales over which a space vehicle could be established in orbit. It should further be noted that far from all of this capacity would be available to SBSP, as other projects have already booked their launches and some lift vehicles might be incompatible with SBSP components.

Energy payback time [days]/[months]:
Energy payback time measures the period it would take, often in days or months, for a given SBSP concept to earn back the same amount of energy that was used to render it operational. This includes energy embedded in materials and expended during production processes as well as the energy necessary to transport all components into orbit. The metric enables us to analyse how long it takes to amortize the energy invested into SBSP generation capacity.

Emissions factor [g CO_{2eq}/kWh]:
Given that SBSP is intended as a renewable energy source, it is important to determine the amount of emissions associated with its energy production and resulting from all stages of its life cycle, including those by some of the required infrastructure, such as launch systems. While we focus on greenhouse gas emissions, the impact can also be extended to all kinds of externalities. For instance, the use of critical raw materials or land use of the rectenna are outside the scope of this thesis but also important when considering the sustainability of SBSP systems.

Based on these four metrics, we now evaluate some of the production and infrastructure necessary to facilitate the large-scale employment of SBSP. We begin with a closer look at manufacturing, assembly, and maintenance dynamics followed by an analysis of the space launch market.

4.2. Manufacturing, assembly, and maintenance

So far, many satellite manufacturing projects have been marred by long delays and large cost overruns. One recent example of this is the James Webb Telescope. Initial cost targets were in the range of $ 1 billion to $ 3.5 billion with the launch slated for 2011 at the latest (Greenfieldboyce, 2021). Ultimately, the space telescope cost more than $ 10 billion dollars and was only launched in 2021. One of the factors driving this disadvantageous dynamic is the lack of economies of scale. Large projects such as the Webb Telescope are complicated one-off satellites with little modularization.

However, scientific satellites' primary goals are not of an economic nature. Conversely, SBSP heavily relies on economies of scale to bring production costs down and accelerate manufacturing processes to ultimately achieve economic competitiveness. Here, modularizing the satellite and rectenna design as much as possible is an important enabler and hence pursued by most concepts as discussed in section 2. Generally, modularity is the idea of dividing a system into smaller elements that can be designed, optimized, and manufactured independently around common measurements and consequently exchanged with other modules in the future (O'Quinn & Jones, 2022).

Recent years have also seen a drastic increase in the amount of commercial satellite projects, which helps fuel cost optimization in space manufacturing through learning curves. Learning curves represent a decrease in specific costs as devices are produced repeatedly and hence the process can be optimized. Typically, the learning factor in the aerospace sector amounts to about 85% (Madonna, 2018). This means that each new batch benefits from a cost reduction of about 15%, reaching a limit once around 50% of initial costs are

reached. The learning factor of 80% assumed for the SPS-ALPHA concept is therefore roughly in line with the industry (Mankins, 2017).

One recent example of strong learning curves as a result of mass-manufacturing modular space vehicles is the solar-powered Starlink system by SpaceX. The company has communicated a per satellite cost for its 4,400-strong initial LEO constellation of well below $ 500,000 at a weight of 275 kg (Mankins, 2021). Calculating an upper bound for specific cost, this results in about 1,800 $/kg, only a fraction of the roughly 300,000 $/kg for standard early satellite types. Most concepts plan to bring hardware costs even lower, with SPS-ALPHA aiming for less than 1,000 $/kg. Regarding manufacturing speed, SpaceX manages to produce about 120 satellites or 33 t per month. This compares to a weight of 2,000 t and a planned construction time, including in-space assembly, for even the lightest SBSP concept of only two years (Way & Lamyman, 2021a).

The difficulty with such modular mass-manufacturing is that it still needs to adhere to high quality and precision standards. Particularly with the RF transmit and receiver elements, uniformity directly translates into system efficiency. For instance, rectenna panels need to show equal DC outputs across all elements to maximise efficiency (Shinohara & Matsumoto, 1998). Furthermore, there is a trade-off between additive manufacturing techniques that could be employed to produce flexible integrated PV and RF sandwich sheets and modular approaches where PV and RF surfaces can still be separated for easier maintenance (Borgue, Panarotto, & Isaksson, 2019).

Related to this, modularisation also proposes great benefits during assembly, scaling up of pilot systems, and maintenance. Given the reliance by all selected concepts on autonomous robotic in-space assembly, having large numbers of identical parts makes this task much easier and allows modules to be added or replaced with relative ease within the overall system constraints. For this purpose, SPS-ALPHA is planning to utilise hexagonal shapes as a common geometry among most of its components (Mankins, 2021). This includes frames, tiles, reflectors and trusses.

Nonetheless, the robots required for such an assembly are arguably also one of the least developed technologies related to SBSP. Given the relevance of such capabilities across the entire space sector, the European Commission has sponsored the MOdular Spacecraft Assembly and Reconfiguration (MOSAR) project to support the development of an autonomous in-space assembly system. After first designs for walking robots have been developed where the two extremities act as legs and arms in an alternating fashion, testing under space conditions is now ongoing (Letier et al., 2019).

Regarding the environmental impact, current assessments show that about 85% of the life cycle impact of a satellite comes from the production of the spacecraft, the launcher, and propellants (Wilson, 2019). Consequently, the technologies and components employed should also be selected for their total impact, including environmental aspects. For instance, there is a large spread in related emissions for different PV technologies discussed in section 2.2. While relatively standard silicone cells tend to have some of the highest resulting emissions, the metrics for newer technologies can vary strongly (Ludin et al., 2018). Pervoskite cells, which we have established as a desirable technology option in chapter 2, can result in anything from 50 to 500 g CO_{2eq}/kWh. This reflects their comparatively low technological maturity and unstable operational lifetime. On the other hand, quantum dot cells display very low emissions at up to 5 g CO_{2eq}/kWh. In concert with other technologies such as nanowires, quantum dots also offer the possibility of reducing embedded emissions further by eliminating the cover glass. Overall, system emissions for SBSP have been estimated to go as low as 20 g CO_{2eq}/kWh (URSI, 2007).

For the end of life, most concepts plan to transfer the satellite into a graveyard orbit as deconstruction would be prohibitively expensive (e.g. Way & Lamyman, 2021a). The exact impacts of moving such a large structure to a graveyard orbit where many other decommissioned satellites can also be found has yet to be determined. Potential collisions could aggravate any of the debris-related issues already present today and create problems for other satellites that are still in service as well as future launches.

Regarding energy payback times, studies in the early 2000s have suggested that these are very competitive compared to terrestrial solar and wind. If only 0.5 GW of SBSP were to be installed, the energy breakeven point including manufacturing would be reached after two years (Summerer & Ongaro, 2005). As the capacity of the system is increased, this quickly falls to below six months, about half as much as other terrestrial renewable technologies. The energy required for transporting the components into orbit plays a significant role in these calculations and is therefore further discussed in the following section.

4.3. Space launch infrastructure

The launch of the components for the space vehicle constitutes a critical step and simultaneously a potential bottleneck for SBSP development. Space launches and the deployment of any asset into orbit have always been a complex and until recently very costly endeavour. Historically, only some of the world's most ambitious and wealthy countries had access to space through government-funded projects. However, over the last decade, closer involvement of the private sector has upended the launch market and fundamentally shifted these dynamics. Today, there is a multitude of commercial players around the globe which are offering or currently developing launch services for different applications. We begin by analysing these changes from a cost and capacity perspective and follow up by some environmental and energy considerations.

4.3.1. Developments in the launch market

The advent of reusable rocket boosters fuelled by NASA's approach to sponsor commercial competition in the launch market has led to a steep decrease in specific launch costs

(Jones, 2018). In figure 12, we have visualised these dynamics by launch vehicle weight class. While no universal definition for these buckets exist, light-lift vehicles correspond to a maximum payload of 2,000 kg to LEO, medium-lift reaches up to 20,000 kg capacity, and heavy-lift vehicles are anything above that (Roberts, 2020). Specific launch costs are then calculated using the mission cost, including all direct and indirect positions, and the maximum payload mass, assuming the rocket is fully expendable and not reused. A configuration for reuse limits the payload mass as additional fuel has to be loaded for the booster to navigate back to earth. Furthermore, it should be noted that for satellites that have their own propulsion system, their mass does not equal launch mass as in-space propellant has to be added.

Looking at the specific launch costs to LEO in figure 12, we can see that the decline is particularly pronounced with medium- and heavy-lift vehicles. One factor in this dynamic is SpaceX. Its Falcon 9 rocket at 2,600 $/kg and Falcon Heavy at 1,500 $/kg are far ahead in terms of price competitiveness compared to any of the other vehicles in operation.

However, launching to geostationary transfer orbit (GTO) instead of LEO as would be required for SBSP changes the equation to some degree. GTO is a transitory orbit that allows the payload to attain GEO with the help of its own on-board propulsion. Consequently, the price per kilogram payload increases. The extent of this change can be shown with some examples. The mission cost for SpaceX's Falcon 9 lies at $62M (SpaceX, 2022). If GTO is targeted instead of LEO, payload capacity decreases by a factor of nearly three from 22,800 kg to 8,300 kg. As this assumes a fully expendable rocket, a mission centred around reuse will bring that number down to 5,500 kg. This would result in specific launch costs of 7,500 $/kg and 11,300 $/kg, respectively. The price increase is even more pronounced with the Falcon Heavy. At $97mio per launch, capacity roughly halves from LEO to GTO to 26,700 kg and mass-specific cost more than doubles to 3,600 $/kg. As the Falcon Heavy configuration includes multiple boosters, having all of them return to earth after depositing the payload lowers these values significantly to a capacity of 8,000 kg at 12,100 $/kg.

In conclusion, GTO commands a price premium by a factor of about two to three above standard launches to LEO, even without reuse. Only looking at launch costs to LEO can therefore be misleading when evaluating the cost performance of SBSP launch infrastructure. Nonetheless, the cost to GTO employing an expendable Falcon Heavy would be below the expectation of 5,000 $/kg by some concepts, such CASSIOPeiA (Cash, 2019).

For the near future, a variety of actors are developing new launch systems which are expected to bring GTO costs down even further. Jeff Bezos' company Blue Origin is looking to transport up to 13,000 kg to GTO (Blue Origin, 2022). To achieve the 5,000 $/kg level, mission costs would have to stay below $70M, which is a reasonable level compared to other vehicles. A joint venture by the defence contractors and aerospace companies Boeing and Lockheed Martin called the United Launch Alliance is also planning to fly its Vulcan rocket for the first time next year (Clark, 2015). The cost target of $100M per mission and a capacity of 13,000 kg to GTO would result in specific costs of about 7,600$/kg.

In comparison, SpaceX's Starship, which has already undergone first flight tests, could drastically undershoot these price levels. Aiming for mission costs of $10M (Roulette, 2022), a 21,000-kg capacity to GTO would result in only about 500 $/kg, a significant decrease from current levels (SpaceX, 2020). Should additional propellant for the launch system be parked in orbit in advance to decrease the amount required on board during launch, an additional decline by a factor of five would be possible as payload capacity to GTO increases to 100 t. NASA also has its Space Launch System in the pipeline, but further specifications have yet to be communicated.

All these new systems would greatly increase the overall capacity of the launch market. Over the last decade, 2,604 t of satellites have been launched into orbit (Euroconsult, 2021). For the next decade, this number is expected to more than double. Nonetheless, more than 550 t per year of total mass launched would compare poorly to what is required to make an entire SBSP system operational. Even if all this capacity were to be used only for the 2,000-t heavy CASSIOPeiA system, it would take four years just to transfer all components into space. For the heaviest concept, MR-SPS, this number would rise to 20. Therefore, launch capacity for now appears to be a serious bottleneck outside the economic considerations for the system.

Furthermore, the lack of common launch or payload interfaces between rockets complicates stowage and load planning, even if all launch systems were available (O'Quinn & Jones, 2022). Despite adaptations towards the commercialisation of satellites, the launch process remains fairly inflexible with periods from booking to actual lift-off reaching two years or more and exact mass properties being required up to 8 months in advance (SpaceX, 2021). With such long lead times and none of the concepts even in the process of manufacturing a prototype yet, the risk remains that launch capacity will be tied up by other projects for the foreseeable future.

In order to utilize the increased payload capacity of launches to LEO and to minimize the need for in-space propellant on board the satellite when transferring to GEO, William Brown has also come up with a concept called Transportronics (Brown, 1992). In essence, the concept would make use of a narrow microwave beam on earth powered by terrestrial renewable energy to fuel a transport device. This transport device would then be able to move satellites from LEO to GEO at low cost, coupled with unprecedented speed for an electronic launch system, and eliminate the need for more costly direct launches to GTO. In the initial study, a system scale up to transfer up to 60,000 t per year was considered as feasible. However, little has come of the idea since.

With the increase of commercial activity in space, a general trend towards smaller and cube satellites has also been observable (O'Quinn & Jones, 2022). This is further reflected

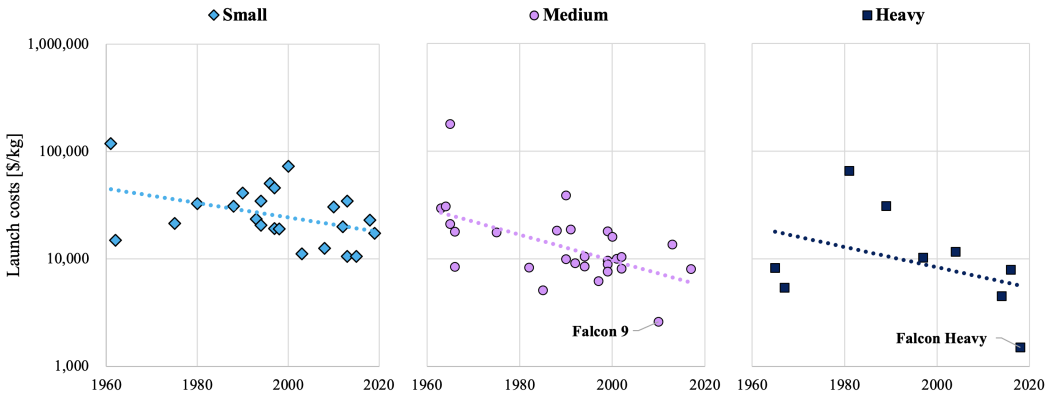


Figure 12: Development of mass-specific launch costs to LEO for different launch vehicle classes with trend lines, measured in 2021 dollars. Own analysis based on data from the Center for Strategic and International Studies (Roberts, 2020)

by the fact that while launched mass is set to double over the next decade, the actual number of satellites put into orbit will more than quadruple (Euroconsult, 2021). As a result, many launch providers are also specializing in smaller satellites, contrary to what a SBSP space vehicle would require. Therefore, modularity is also critical to make SBSP compatible with as many launch systems as possible.

4.3.2. Environmental and energy payback considerations

The environmental impact of rocket launches strongly correlates with the type of fuel that is used. In the past, solid fuels were the dominant variant (Dallas, Raval, Gaitan, Saydam, & Dempster, 2020). However, due to their ozone-depleting and toxic characteristics, which were particularly observable in Kazakhstan as the former base for Soviet launches, the launch market is moving to alternatives. Looking forward, liquid hydrogen (LH2) and oxygen (LOx) as well as kerosene are likely to become more dominant. While LOx/LH2 fuels are difficult to handle due to the temperatures required, they provide great energy output and only result in water vapour when burnt.

Nonetheless, the entire SBSP system and the launch infrastructure in particular will require a full life cycle analysis (LCA). First analyses based on a framework developed for space missions by Wilson (2019) show that the most significant environmental impacts of space launches include climate impacts, toxicity, acidification, and debris. In particular, the potential to cause ozone depletion, freshwater aquatic ecotoxicity, and air acidification are the prominent three effects. Based on the numbers compiled by Wilson (2019) for 2018, the emissions impact of a single space mission is more than 1,000 times greater than that of the average global flight. Given the expected expansion in the sector, close attention will therefore have to be paid to these environmental effects and options that minimise negative externalities. For instance, LOx/LH2 fuels where the hydrogen is generated via green electricity promise to be one of those sustainable alternatives.

The fact that hydrogen can be produced via electrolyses also offers the potential for SBSP to recoup the fuel that was used to put the system into space directly. It roughly takes 32 kWh in propellant to move one kilogram from the earth's surface to GEO (Mankins, 2017). Based on that metric, it would take CASSIOPeiA just over one day, SPS-ALPHA five days, and MR-SPS one and a half weeks to generate the same amount of energy that was used in the form of fuel. If the systems were to produce their own LH2-based propellant, for instance during times where terrestrial solar and wind on their own are sufficient, we would also have to account for the energy lost during conversion steps and required for liquefaction. Consequently, the payback time would increase to about three weeks for CASSIOpeiA, three months for SPS-ALPHA, and half a year for the comparatively heavy MR-SPS. A summary of these energy payback times is also provided in table 4.

4.4. Conclusion

Overall, a modular approach to any SBSP concept will be critical to enable economies of scale during manufacturing, compatibility with as many launch systems as possible given the recent trend towards smaller satellites, and ease of assembly as well as maintenance. The Starlink project has successfully shown that manufacturing at comparatively low mass-specific cost is possible already. While the amount of Starlink modules only number in the low thousands, SBSP concepts would far surpass that amount, potentially unlocking greater learning curves.

The continuous decline in launch costs driven by the commercialisation of the launch market also promises to further enhance the economics of SBSP. Based on some of the variance analysis for LCOE in the literature (e.g. Way & Lamyman, 2021a), launch costs are even a particularly strong lever. However, our analysis suggests that total payload capacity will lag behind what would likely be required to transfer an entire SBSP system to space for the foreseeable future. Upcoming heavy-lift vehicles such as Vulcan, Starship, and SLS might be able to alleviate this bottleneck to some degree. Nonetheless, their exact market entry is as of yet unknown. Additionally, the long lead times in the market could make it difficult to achieve any meaningful presence in orbit over

Table 4: The amount of time it would take for different SBSP concepts to feed the amount of energy required to put them into space back into the grid. The first number is a simplified comparison, only considering the energy content of the fuel used for a launch to GEO. The second number also accounts for all conversion steps necessary to produce green LH2, using the electricity generated by SBSP. Own analysis based on data from Cash (2021a) and Mankins (2017).

		Energy payback time [days]	
		Energy content of fuel needed for launch to GEO	Energy needed to produce green LH2 propellant for launch to GEO
Concepts	SPS-ALPHA	5.0	91.7
	CASSIOPeiA	1.3	23.0
	MR-SPS	13.3	242.4

the next couple of years. Therefore, the commonly in the literature observable singular focus on launch costs might be misguided as the true constraint potentially lies in the availability of payload space.

Environmentally and from an energy perspective, the literature suggests that SBSP at scale has shorter energy payback times than conventional renewables at relatively low emissions. Furthermore, a shift towards more sustainable rocket fuels could help lower emissions by using green energy to produce the liquid hydrogen utilized by the launch vehicles. Nonetheless, given the majority of impacts arise during the manufacturing process of the mission, particular focus will have to be placed on using sustainable materials and processes. So far, it seems that some of the technologies introduced in section 2.2 such as quantum dots could meet this requirement. At the end, a holistic LCA will be necessary to fully understand all the impacts of SBSP. Particularly Wilson (2019) has developed a framework that could be applicable to such an analysis of SBSP.

5. Summary and outlook

SBSP combines two of the most dynamic industries of our time: renewable energy and space. It is unique in its potential ability to offer flexible renewable baseload power that is dispatchable on a semi-global scale. Such capabilities would constitute a crucial step in securing energy security for the electrification of our economy. There are already a number of concepts that have been far advanced in their planning stages, often with direct government involvement and support. Nonetheless, none of them have produced a full-system prototype yet.

For our analysis, we have divided the general SBSP model into three segments: the space segment, wireless power transmission (WPT) and ground structures, and manufacturing and infrastructure. Based on the metrics defined for each of those, we have seen that some factors remain which are still holding SBSP back from becoming a reality. When it comes to solving these issues, we are often faced with tradeoffs between different system parameters. The right solution will ultimately have to be found through prototyping and demonstrations, which should be the priority for any SBSP design going forward.

One of the biggest levers to improve the economics of the space segment are weight and power. Hence, mass-specific power and aerial density were identified as crucial determinants of space segment performance. Out of the three investigated concepts, CASSIOPeiA achieved the highest mass-specific power by a notable margin at 1.51 W/g. These results indicate that the consequent use of sandwich modules and boosting of PV power through solar concentration are essential in achieving competitive performance levels. This was particularly evident with the Chinese MR-SPS concept, where the lack of PV and RF surface integration as well as non-existent solar concentration led to some of the poorest results at 0.16 W/g. However, SPS-ALPHA also shows that sandwich modules as well as any reflector arrays and structures have to be lightweight so their benefits are not neutralised by the additional mass.

Still, even for CASSIOPeiA there appears to be room for improvement. We have been able to identify a number of technology alternatives that promise to reduce weight, improve power levels, and enhance inherent radiation resistance. Some approaches could even deliver multiple of these benefits at once. For instance, LSCs bear the potential to increase PV cell performance while maintaining a flat-plate tile geometry and even boosting radiation resistance. Nanowires could complement any design to drastically reduce radiation damage and increase the useful lifespan of the space segment while eliminating the cover glass. Making such protective but weighty components redundant could boost mass-specific power by a factor of nearly three. Lastly, pervoskite as a cell material has the potential for significant synergies with the aforementioned technologies, based on its self-healing properties and exceptionally high power output. However, uncertainties regarding the efficiency levels of all three of these alternatives, particularly when under high thermal stress, will have to be addressed before they can be fully implemented. Nonetheless, even based on their current performance levels they appear competitive, suggesting that foregoing some efficiency to significantly reduce mass might be an approach worth pursuing.

On the other hand, not all investigated options appear to work for SBSP systems. Despite having by far the highest mass-specific power at more than 2 W/g, integrated

parabolic mirrors appear to be an infeasible solution as they are difficult to mass-manufacture and notably constrain satellite architecture. But even with the flat-plate module design maintained, the challenge of achieving a continuous duty cycle remains. Our two proposed solutions of dual-sided sandwich modules for planar arrays or L-shaped modules for helical arrays both come with inherent trade-offs. While the former results in an increase in complexity, weight, and hence cost, the latter requires heavy reflector structures and has limited scalability. Ultimately, both designs might be viable within the right setting. However, planar arrays would still require extensive attitude control and adjustment maneuvers, even when employing dual-sided tiles.

For the transmission segment of the system, efficiencies are a key determinant of subsystem performance as they dictate the amount of energy that can ultimately be fed into the grid. Given the need for atmospheric and weather penetration as well as scalability, microwaves appear to be the modality of choice, beating out lasers, who particularly struggle when it comes to those two characteristics. We have shown that there is an extensive history of successful MPT demonstrations, achieving and even surpassing the link efficiencies of above 50% targeted by today's most advanced concepts, such as CASSIOPeiA or SPS-ALPHA. The prevailing notion based on far-field assumptions that MPT cannot achieve the performance levels necessary for SBSP therefore appears to be unfounded.

Nonetheless, the distances across which power has been beamed and the corresponding scale of transmit and receiver apertures still lags behind what would be required for a space-to-earth system. This raises the questions of whether the success of terrestrial demonstrations will ultimately be transferable to the space environment. Ongoing and upcoming WPT experiments should help in overcoming these uncertainties, as the technology enjoys considerable momentum due to its applicability outside of SBSP.

The biggest remaining complication appears to be the question of spectrum management. With no regulatory definition of WPT and hence no internationally allocated bandwidths, there is no common frequency around which systems could be optimised and advanced. As a result, each actor is left to define their own standards. Japan, a notable contributor when it comes to starting the necessary processes at the ITU, should therefore be supported in its efforts by other countries to the benefit of everyone pursuing this technology.

Lastly, we analysed the manufacture and infrastructure subsystem with a particular focus on the launch market. For the construction of the satellite and ground structure, economies of scale through modularisation appear crucial to unlock cost benefits. The example of Starlink has shown how far-reaching these advantages can be. Nonetheless, while a first analysis suggests energy payback times and environmental impacts from the manufacturing process are manageable, in-depth environmental studies and LCAs are required to ensure the overall sustainability of the technology. Furthermore, many approaches to SBSP rely on autonomous robotic assembly and maintenance once in space. However, this technology is as of yet untested in a practical setting, despite being so crucial to the success of these concepts.

When examining the transfer from earth to orbit, launch costs are often cited in the literature as a primary concern and obstacle for the realisation of an economically viable SBSP system. Fortunately, the commercialisation of the launch market over the last decade has brought prices down by orders of magnitude while capacities are still expanding, enhancing SBSP economics in turn. Furthermore, greener fuel alternatives such as propellants based on LOx/LH2 offer the possibility of significantly reducing some of the environmental issues that have raised concerns about rocket launches in the past.

Nonetheless, it appears questionable whether the 550 t of planned total yearly launch capacity over the next decade will be enough. Given that most of that capacity would not be available on an exclusive basis, even the lightest concept CASSIOPeiA could arguably not be transferred into orbit within a reasonable timeframe. Additionally, the long lead times for launch missions plus a lack of common payload interfaces make it very difficult to flexibly spread out space transfer operations across multiple different lift vehicles. The recent trend towards smaller satellites, which is also reflected in the reduced payload capacities of many upcoming new launch systems, further underlines the need for a modular approach to ensure compatibility with these small-scale lift vehicles. Consequently, our analysis suggests that launch capacity rather than cost might be the determining aspect concerning the operationality of SBSP. A reduction in mass would then not only be a matter of economics but might be the factor that would render such a project possible in the first place.

In conclusion, we were able to identify a number of technology alternatives that have the potential of improving critical subsystem metrics to ultimately enhance SBSP system economics. These could lead to SBSP being an overall competitive renewable energy alternative to drive wide-spread electrification of industries. The exact magnitude of these economic benefits will have to be confirmed through dedicated LCOE studies. Furthermore, our analysis based on past demonstrations suggests that WPT is already at a stage where, if scaled up, could deliver the required transmission capabilities. However, the launch infrastructure appears to be the critical bottleneck that could prevent any SBSP system from advancing past the planning stage in the foreseeable future. If these capacity constraints are not addressed, the future of SBSP will likely remain uncertain.

References

Abdelal, G. F., Gad, A. H., & Abuelfoutouh, N. (2013). *Finite element analysis for satellite structures - applications to their design, manufacture and testing*. Springer.

Akiba, R., Miura, K., Hinada, M., Matsumoto, H., & Kaya, N. (1993). ISY-METS rocket experiment. *The Institute of Space and Astronautical Science report*, *652*, 1–13.

Andreev, V. M. (2018). GaAs and high-efficiency space cells. In *Mcevoy's handbook of photovoltaics* (pp. 421–438). Elsevier.

Barrigón, E., Heurlin, M., Bi, Z., Monemar, B., & Samuelson, L. (2019). Synthesis and applications of III–V nanowires. *Chemical reviews*, *119*(15), 9170–9220.

Blue Origin. (2022, April). *New Glenn capabilities.* Retrieved 10 March 2021, from https://www.blueorigin.com/new-glenn/

Borgue, O., Panarotto, M., & Isaksson, O. (2019). Modular product design for additive manufacturing of satellite components: maximising product value using genetic algorithms. *Concurrent Engineering*, *27*(4), 331–346.

Brown, W. C. (1965, December). *Experimental airborne microwave supported platform* (Contract AF30 (602) 3481 No. RADC-TR-65-188). NASA.

Brown, W. C. (1974). The technology and application of free-space power transmission by microwave beam. *Proceedings of the IEEE*, *62*(1), 11–25.

Brown, W. C. (1977). Electronic and mechanical improvement of the receiving terminal of a free-space microwave power transmission system. *NASA STI/Recon Technical Report*, *77*, 31613.

Brown, W. C. (1980a). The history of the development of the rectenna. In *Solar power satellite microwave power transmission and reception* (p. 271).

Brown, W. C. (1980b). Solar power satellite program rev. DOE/NASA satellite power system concept development evaluation program. In *Final proc. conf* (Vol. 800491, pp. 8–18).

Brown, W. C. (1992). *A transportronic solution to the problem of interorbital transportation* (Tech. Rep.).

Brown, W. C., & Eves, E. E. (1992). Beamed microwave power transmission and its application to space. *IEEE Transactions on Microwave Theory and Techniques*, *40*(6), 1239–1250.

Brown, W. C., George, R. H., Heenan, N. I., & Wonson, R. C. (1969, May 05). *Microwave to DC converter.* Google Patents. (US Patent 3434678A)

Brown, W. C., & Triner, J. (1982). Experimental thin-film, etched-circuit rectenna. In *1982 IEEE MTT-S International Microwave Symposium Digest* (pp. 185–187).

Cash, I. (2017). CASSIOPeiA solar power satellite. In *Proc. 2017 IEEE International Conference on Wireless for Space and Extreme Environments (WiSEE)* (pp. 144–149).

Cash, I. (2019). CASSIOPeiA – a new paradigm for space solar power. *Acta Astronautica*, *159*, 170–178.

Cash, I. (2021a, 09 December). *CASSIOPeiA SPS – space based solar power for net zero.* Workshop: ESA's 1st International Workshop for Space Solar Power for Net Zero.

Cash, I. (2021b). *A phased array antenna and apparatus incorporating the same.* Intellectual Property Office UK. (UK Patent GB2563574)

Cavalli, A., Dijkstra, A., Haverkort, J. E., & Bakkers, E. P. (2018). Nanowire polymer transfer for enhanced solar cell performance and lower cost. *Nano-Structures & Nano-Objects*, *16*, 59-62.

Chen, Q. (2020). *Research on high-performance receiving and rectifying technology for microwave wireless power transmission* (Unpublished doctoral dissertation). Sichuan University.

Cheney, M. (1981). *Tesla, man out of time*. Englewood Cliffs, NJ, USA: Prentice-Hall.

Clark, S. (2015, April 22). *ULA needs commercial customers to close Vulcan rocket business case.* Retrieved 25 January 2022, from https://spaceflightnow.com/2015/04/22/ula-needs-commercial-business-to-close-vulcan-rocket-business-case/

Corson, D., Bradburn, N., Chodorow, M., Dougherty, J., Gordon, W., Hannon, B., … others (1981). Electric power from orbit: A critique of a satellite power system. *National Academy of Sciences, Washington, DC*.

Cui, Y., Duan, X., Hu, J., & Lieber, C. M. (2000). Doping and electrical transport in silicon nanowires. *The Journal of Physical Chemistry B*, *104*(22), 5213–5216.

Dallas, J., Raval, S., Gaitan, J. A., Saydam, S., & Dempster, A. (2020). The environmental impact of emissions from space launches: A comprehensive review. *Journal of Cleaner Production*, *255*, 120209.

Davis, H. P. (2012). *Space-based solar power, an update* (Tech. Rep.). Solar High Study Group.

Dickinson, R. M. (1975). Evaluation of a microwave high-power reception-conversion array for wireless power transmission. *NASA STI/Recon Technical Report*, *76*, 11207.

Dickinson, R. M. (2003). Wireless power transmission technology state of the art – the first bill brown lecture. *Acta Astronautica*, *53*(4-10), 561–570.

Dickinson, R. M., & Brown, W. C. (1975). Radiated microwave power transmission system efficiency measurements. *NASA Tech. Memo 33-727*.

Dickinson, R. M., & Maynard, O. (1999). Ground based wireless and wired power transmission cost comparison. *JPL TRS 1992+*.

DoC, & NTIA. (2021, January). Manual of regulations and procedures for federal radio frequency management. *NTIA Manual*.

Dong, Y., Dong, S.-W., Wang, Y., Liu, S., Li, X., Gao, S., … Ran, L. (2018). Focused microwave power transmission system with high-efficiency rectifying surface. *IET Microwaves, Antennas & Propagation*, *12*(5), 808–813.

Espinet-Gonzalez, P., Barrigón, E., Chen, Y., Otnes, G., Vescovi, G., Mann, C., … others (2020). Nanowire solar cells: a new radiation hard PV technology for space applications. *IEEE Journal of Photovoltaics*, *10*(2), 502–507.

Espinet-Gonzalez, P., Barrigón, E., Otnes, G., Vescovi, G., Mann, C., France, R. M., … others (2019). Radiation tolerant nanowire array solar cells. *ACS Nano*, *13*(11), 12860–12869.

Etani, S., Iwashita, M., & Kaya, N. (2011). Development on the sandwich panel for the practical solar power satellite. In *Proc. of the 28th international symposium on space technology and science.*

Euroconsult. (2021, December). *Satellites to be built and launched by 2030* (No. 24).

Foust, J. (2008, 15 September). A step forward for space solar power. *The Space Review*. Retrieved 13 April 2022, from https://www.thespacereview.com/article/1210/1

Fujino, Y., et al. (1993, February). A rectenna for MILAX. In *Proc. wireless power transmissions conference* (pp. 273–277).

Gal-Katziri, M., Fikes, A., Bohn, F., Abiri, B., Hashemi, M. R., & Hajimiri, A. (2020). Scalable, deployable, flexible phased array sheets. In *Proc. 2020 IEEE/MTT-S international microwave symposium (IMS)* (pp. 1085–1088).

Gal-Katziri, M., & Hajimiri, A. (2018). A sub-picosecond hybrid dll for large-scale phased array synchronization. In *2018 IEEE Asian solid-state circuits conference (A-SSCC)* (pp. 231–234).

Gal-Katziri, M., Ives, C., Khakpour, A., & Hajimiri, A. (2022). Optically synchronized phased arrays in CMOS. *IEEE Journal of Solid-State Circuits*.

Garcia, M. (2021, November 4). *International space station facts and figures.* NASA. Retrieved 09 January 2022, from https://www.nasa.gov/feature/facts-and-figures

Gdoutos, E., Leclerc, C., Royer, F., Türk, D. A., & Pellegrino, S. (2019). Ultralight spacecraft structure prototype. In *AIAA Scitech 2019 Forum* (p. 1749).

Gdoutos, E., Truong, A., Pedivellano, A., Royer, F., & Pellegrino, S. (2020). Ultralight deployable space structure prototype. In *AIAA Scitech 2020 Forum* (p. 0692).

George, R., & Okress, E. (1968). Solid state power rectifications. *Microwave power engineering*, *1*, 275–294.

Gibb, J. (2018). Lightweight flexible space solar arrays, past, present and future. In *2018 IEEE 7th world conference on photovoltaic energy conversion (WCPEC) (A joint conference of 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC)* (p. 3530-3534).

Glaser, P. (1968). Power from the sun: Its future. *Science*, *162*(3856), 857–861.

Glaser, P. (1973, December 25). *Method and apparatus for converting solar radiation to electrical power.* Google Patents. (US Patent 3,781,647)

Goel, A., Chung, S.-J., & Pellegrino, S. (2017, June). Trajectory design of a spacecraft formation for space-based solar power using sequential convex programming. In *Proc. conference 9th international workshop on satellite constellations and formation flying (IWSCFF)*.

Goel, A., Lee, N., & Pellegrino, S. (2017). Trajectory design of formation flying constellation for space-based solar power. In *2017 IEEE aerospace conference* (pp. 1–11).

Grandidier, J., Jaffe, P., Roberts, W. T., Wright, M. W., Fraeman, A. A., Raymond, C. A., ... others (2021). Laser power beaming for lunar night and permanently shadowed regions. *Jet Propulsion*, *818*, 354–1566.

Grebennikov, A. (2011). *RF and microwave transmitter design* (Vol. 234). John Wiley & Sons.

Greenfieldboyce, N. (2021, December 22). *Why some astronomers once feared NASA's James Webb space telescope would never launch.* Retrieved 04 April 2022, from https://n.pr/3vgyenz

Hagerty, J. A., & Popovic, Z. (2001). An experimental and theoretical characterization of a broadband arbitrarily-polarized rectenna array. In *2001 IEEE MTT-S international microwave symposium digest (Cat. No. 01CH37157)* (Vol. 3, pp. 1855–1858).

Hajimiri, A., Abiri, B., Bohn, F., Gal-Katziri, M., & Manohara, M. H. (2020). Dynamic focusing of large arrays for wireless power transfer and beyond. *IEEE Journal of Solid-State Circuits*.

Hajimiri, S. A., Atwater, H. A., Pellegrino, S., Abiri, B., & Bohn, F. (2021). *Large-scale space-based solar power station: power transmission using steerable beams.* Google Patents. (US Patent 11,128,179)

Hashemi, M. R. M., Fikes, A. C., Gal-Katziri, M., Abiri, B., Bohn, F., Safaripour, A., ... others (2019). A flexible phased array system with low areal mass density. *Nature Electronics*, *2*(5), 195–205.

Hawkins, J., Houston, S., Hatfield, M., & Brown, W. (1998). The saber microwave-powered helicopter project and related wpt research at the University of Alaska Fairbanks. In *AIP conference proceedings* (Vol. 420, pp. 1092–1097).

Heinzler, R., Schindler, P., Seekircher, J., Ritter, W., & Stork, W. (2019). Weather influence and classification with automotive lidar sensors. In *2019 IEEE intelligent vehicles symposium (IV)* (pp. 1527–1534).

Hou, X., & Li, M. (2021, 09 December). *Activities of SPS development in China.* Workshop: ESA's 1st International Workshop for Space Solar Power for Net Zero.

Hu, L., Zhao, Q., Huang, S., Zheng, J., Guan, X., Patterson, R., ... others (2021). Flexible and efficient perovskite quantum dot solar cells via hybrid interfacial architecture. *Nature Communications*, *12*(1), 1–9.

IEA. (2020). *World energy outlook 2020*. OECD Publishing.

IEEE Standards Coordinating Committee. (2019, October). *IEEE standard for safety levels with respect to human exposure to radio frequency electromagnetic fields, 3kHz to 300GHz* (Tech. Rep. No. IEEE Std C95.1-2019).

IPCC. (2022, April). *AR6 Synthesis Report: Climate Change 2022.* Retrieved 10 April 2022, from https://www.ipcc.ch/assessment-report/ar6/

ITU. (2019, August). *Attenuation due to clouds and fog* (Standard No. ITU-R P.840). International Telecommunication Union (ITU).

ITU. (2021a, 06). *Applications of wireless power transmission via radio frequency beam* (Tech. Rep. No. ITU-R SM.2392-1). International Telecommunication Union (ITU).

ITU. (2021b, October). *ITU-R: Managing the radio-frequency spectrum for the world.* Retrieved 25 January 2022, from https://www.itu.int/en/mediacentre/backgrounders/Pages/itu-r-managing-the-radio-frequency-spectrum-for-the-world.aspx

Jaffe, P. (2013). *A sunlight to microwave power transmission module prototype for space solar power* (Unpublished doctoral dissertation). University of Maryland, College Park.

Jaffe, P. (2020). 24 - Space Solar. In T. M. Letcher (Ed.), *Future energy* (Third ed., p. 519-542). Elsevier.

Jehn, R., Agapov, V., & Hernández, C. (2005). The situation in the geostationary ring. *Advances in space research*, *35*(7), 1318–1327.

Jones, H. (2018, July). The recent large reduction in space launch cost. In *Proc. 48th international conference on environmental systems.*

Kalyuzhnyy, N. A., Emelyanov, V. M., Evstropov, V. V., Mintairov, S. A., Mintairov, M. A., Nahimovich, M. V., ... Shvarts, M. Z. (2020). Optimization of photoelectric parameters of InGaAs metamorphic laser (λ= 1064 nm) power converters with over 50% efficiency. *Solar Energy Materials and Solar Cells*, *217*, 110710.

Kang, S., Jeong, J., Cho, S., Yoon, Y. J., Park, S., Lim, S., ... Ko, H. (2019). Ultrathin, lightweight and flexible perovskite solar cells with an excellent power-per-weight performance. *Journal of Materials Chemistry A*, *7*(3), 1107–1114.

Kawashima, N., & Takeda, K. (2004). 1.2 km laser energy transmission for the development of a lunar rover confirming the presence of ice on the moon (AAS 03-737). In *International lunar conference 2003* (Vol. 108, p. 291).

Kawashima, N., & Takeda, K. (2008). Laser energy transmission for a wireless energy supply to robots. *Robotics and Automation in Construction*, *10*, 373–380.

Kaya, N. (1996). Transmitting antenna system for airship demonstration (ETHER). *Space Energy and Transportation*, *1*(4), 237–245.

Kaya, N., Kojima, H., Matsumoto, H., Hinada, M., & Akiba, R. (1994). ISY-METS rocket experiment for microwave energy transmission. *Acta Astronautica*, *34*, 43-46.

Kayes, B. M., Zhang, L., Twist, R., Ding, I.-K., & Higashi, G. S. (2014). Flexible thin-film tandem solar cells with gt;30efficiency. *IEEE Journal of Photovoltaics*, *4*(2), 729-733.

Kazmierski, T. J., & Beeby, S. (2014). *Energy harvesting systems*. Springer.

Kelzenberg, M. D., Espinet-Gonzalez, P., Vaidya, N., Warmann, E. C., Naqavi, A., Loke, S. P., ... others (2018). Ultralight energy converter tile for the space solar power initiative. In *2018 IEEE 7th world conference on photovoltaic energy conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)* (pp. 3357–3359).

Kim, N.-K., Min, Y. H., Noh, S., Cho, E., Jeong, G., Joo, M., ... others (2017). Investigation of thermally induced degradation in CH3NH3PbI3 perovskite solar cells using in-situ synchrotron radiation analysis. *Scientific reports*, *7*(1), 1–9.

Koblin, W., Krüger, E., & Schuh, U. (1984). *Handbuch Passive Nutzung der Sonnenenergie*. Bonn: Ministerium für Raumordnung, Bauwesen und Städtebau.

Koomanoff, F. A. (1981). Satellite power system concept development and evaluation program. *Space Sol. Power Rev.*(CONF-800672-2).

Lang, F., Nickel, N. H., Bundesmann, J., Seidel, S., Denker, A., Albrecht, S., ... others (2016). Radiation hardness and self-healing of perovskite solar cells. *Advanced Materials*, *28*(39), 8726–8731.

Larson, W. J., & Wertz, J. R. (1992). *Space mission analysis and design* (Tech. Rep.). Torrance, CA (United States); Microcosm, Inc.

Law, D. C., Edmondson, K., Siddiqi, N., Paredes, A., King, R., Glenn, G., ... Karam, N. (2006). Lightweight, flexible, high-efficiency III-V multijunction cells. In *2006 IEEE 4th world conference on photovoltaic energy conference* (Vol. 2, p. 1879-1882).

Lei, Z. (2019, 27 February). *Scientists envision solar power station in space.* Retrieved 04 March 2022, from http://www.chinadaily.com.cn/a/201902/27/WS5c75c8b3a3106c65c34eb8e3.html

Letier, P., Yan, X. T., Deremetz, M., Bianco, A., Grunwald, G., Roa, M., ... others (2019). MOSAR: Modular spacecraft assembly and reconfiguration demonstrator. In *Proc. 15th symposium on advanced space technologies in robotics and automation.*

Ludin, N. A., Mustafa, N. I., Hanafiah, M. M., Ibrahim, M. A., Teridi, M. A. M., Sepeai, S., ... Sopian, K. (2018). Prospects of life cycle assessment of renewable energy from solar photovoltaic technologies: a review. *Renewable and Sustainable Energy Reviews*, *96*, 11–28.

Madonna, R. G. (2018). Use of an iterative research and development–system engineering approach for the caltech space solar power project. In *Proc. 2018 6th IEEE international conference on wireless for space and extreme environments (WiSEE)* (pp. 200–205).

Madonna, R. G. (2021, 09 December). *Caltech space solar power project.* Workshop: ESA's 1st International Workshop for Space Solar Power for Net Zero.

Malinkiewicz, O., Imaizumi, M., Sapkota, S. B., Ohshima, T., & Öz, S. (2020). Radiation effects on the performance of flexible perovskite solar cells for space applications. *Emergent Materials*, *3*(1), 9–14.

Mankins, J. (1997). A fresh look at space solar power: New architectures, concepts and technologies. *Acta Astronautica*, *41*(4-10), 347–359.

Mankins, J. (2017). New developments in space solar power. *NSS Space Settlement Journal*, 1–30.

Mankins, J. (2021, 09 December). *SPS-ALPHA and the economics of space solar power.* Workshop: ESA's 1st International Workshop for Space Solar Power for Net Zero.

Mankins, J., Kaya, N., & Vasile, M. (2012). SPS-ALPHA: The first practical solar power satellite via arbitrarily large phased array (a 2011-2012 NIAC project). In *Proc. 10th international energy conversion engineer-*

ing conference (p. 3978).

Marshall, M. A., Goel, A., & Pellegrino, S. (2020). Power-optimal guidance for planar space solar power satellites. *Journal of Guidance, Control, and Dynamics*, *43*(3), 518–535.

Marshall, M. A., Madonna, R. G., & Pellegrino, S. (2021). Investigation of equatorial medium earth orbits for space solar power. *IEEE Transactions on Aerospace and Electronic Systems*.

Masrur, M. A., & Cox, M. (2019). A unique military application of wireless power transfer: Wireless charging through a vehicle seat with simplified design considerations. *IEEE Industrial Electronics Magazine*, *13*(4), 19–30.

Matsumoto, H. (1989). Microwave energy transmission. *Nippon Koku Uchu Gakkaishi (Japan)*, *37*(422).

Matsumoto, H. (2002). Research on solar power satellites and microwave power transmission in Japan. *IEEE microwave magazine*, *3*(4), 36–45.

Matsumoto, H., Kaya, N., Kimura, I., Miyatake, S., Nagatomo, M., & Obayashi, T. (1982). MINIX project toward the solar power satellite-rocket experiment of microwave energy transmission and associated nonlinear plasma physics in the ionosphere. In *Proc. ISAS space energy symposium* (pp. 69–76).

Maynard, O. E., & Blick, E. F. (1980). *Solid state SPS microwave generation and transmission study* (Vol. 3338). NASA.

McSpadden, J. O., & Mankins, J. C. (2002). Space solar power programs and microwave wireless power transmission technology. *IEEE microwave magazine*, *3*(4), 46–57.

Mihara, S., Maekawa, K., Nakamura, S., Sasaki, K., Homma, Y., Hangai, M., … others (2018). The plan of microwave power transmission development for SSPS and its industry application. In *Proc. 2018 asia-pacific microwave conference (APMC)* (pp. 443–445).

Mihara, S., Sato, M., Nakamura, S., Sasaki, K., Homma, Y., Sasaki, T., … Fujiwara, T. (2015). The result of ground experiment of microwave wireless power transmission. In *Proc. 66th international astronautic congress*.

Miyazawa, Y., Ikegami, M., Chen, H.-W., Ohshima, T., Imaizumi, M., Hirose, K., & Miyasaka, T. (2018). Tolerance of perovskite solar cell to high-energy particle irradiations in space environment. *iScience*, *2*, 148–155.

Mizojiri, S., Takagi, K., Shimamura, K., Yokota, S., Fukunari, M., Tatematsu, Y., & Saito, T. (2019). Demonstration of sub-terahertz coplanar rectenna using 265 GHz gyrotron. In *Proc. 2019 IEEE wireless power transfer conference (WPTC)* (pp. 409–412).

NASA. (1978, October). *Satellite power system concept development and evaluation program, reference system report* (Tech. Rep. No. Rep. DOE/ER-0023). Department of Energy.

Needell, D. R., Bauser, H., Phelan, M., Bukowsky, C. R., Ilic, O., Kelzenberg, M. D., & Atwater, H. A. (2019). Ultralight luminescent solar concentrators for space solar power systems. In *Proc. 2019 IEEE 46th photovoltaic specialists conference (PVSC)* (pp. 2798–2801).

Needell, D. R., Ilic, O., Bukowsky, C. R., Nett, Z., Xu, L., He, J., … others (2018). Design criteria for micro-optical tandem luminescent solar concentrators. *IEEE Journal of Photovoltaics*, *8*(6), 1560–1567.

Needell, D. R., Nett, Z., Ilic, O., Bukowsky, C. R., He, J., Xu, L., … others (2017). Micro-optical tandem luminescent solar concentrator. In *Proc. 2017 IEEE 44th photovoltaic specialist conference (PVSC)* (pp. 1737–1740).

Nishioka, T., & Yano, S. (2015, 16 March). *Mitsubishi heavy takes step toward long-distance wireless power.* Retrieved 20 December 2021, from https://asia.nikkei.com/Business/Biotechnology/Mitsubishi-Heavy-takes-step-toward-long-distance-wireless-power

NRL. (2019, October). *Press release: Researchers transmit energy with laser in 'historic' power-beaming demonstration.* Retrieved 13 February 2022, from https://bit.ly/37M5QBw

NRL. (2020, 18 May). *Press release: Nrl conducts first test of solar power satellite hardware in orbit.* Retrieved 13 February 2022, from https://bit.ly/3vfPIR4

Nugent, T., Bashford, D., Bashford, T., Sayles, T., & Hay, A. (2020). Long-range integrated safe laser power beaming demonstration. In *Proc. optical wireless fiber power transmission conference (OWPT)* (pp. 12–13).

O'Quinn, C., & Jones, K. (2022). Increased access to space with modularity and interface standards. In *AIAA Scitech 2022 forum* (p. 647).

Park, Y., Kim, K., & Youn, D. (2018). Rectenna array design for receiving high power in beam type wireless power transmission. In *Proc. 2018 Asia-Pacific microwave conference (APMC)* (pp. 440–442).

Pedivellano, A., Gdoutos, E., & Pellegrino, S. (2020). Sequentially controlled dynamic deployment of ultra-thin shell structures. In *AIAA Scitech 2020 forum*. American Institute of Aeronautics and Astronautics.

Pellegrino, S., Atwater, H. A., Hajimiri, S. A., Gdoutos, E. E., Leclerc, C., Royer, F. A., & Pedivellano, A. (2020, January 23). *Coilable thin-walled longerons and coilable structures implementing longerons and methods for their manufacture and coiling.* Google Patents. (US Patent App. 16/514,793)

Penn, J., Law, G., et al. (2001). The aerospace corporation systems studies and analysis of the space solar power (SSP) exploratory research and technologies concepts and applications. *Aerospace Report No. ATR-01 (7710)-1*, *9*, 106.

Philipps, S. P., Dimroth, F., & Bett, A. W. (2018). High-efficiency III–V multijunction solar cells. In *McEvoy's handbook of photovoltaics* (pp. 439–472). Elsevier.

Pisacane, V. L. (2005). *Fundamentals of space systems*. Johns Hopkins University Applied Physics Laboratory Series in Science and Engineering.

Potter, P. D. (1961). A new horn antenna with suppressed sidelobes and equal beam widths. *Microwave journal*, *6*(6), 71–78.

Qiu, H., Liu, H., Jia, X., Jiang, Z.-Y., Liu, Y.-H., Xu, J., … Chen, K. J. (2021). Compact, flexible, and transparent antennas based on embedded metallic mesh for wearable devices in 5G wireless network. *IEEE Transactions on Antennas and Propagation*, *69*(4), 1864–1873.

Roberts, T. (2020, 02 September). *Space launch to low earth orbit: How much does it cost?* Retrieved 26 March 2022, from https://bit.ly/3rojZMv

Rodenbeck, C. T., Jaffe, P. I., Strassner II, B. H., Hausgen, P. E., McSpadden, J. O., Kazemi, H., … Self, A. P. (2021). Microwave and millimeter wave power beaming. *IEEE Journal of Microwaves*, *1*(1), 229-259.

Rodenbeck, C. T., Tierney, B. B., Park, J., Parent, M. G., DePuma, C. B., Bauder, C. J., … Mayhan, T. (2022). Terrestrial microwave power beaming. *IEEE Journal of Microwaves*.

Rogers, J. E., & Spirnak, G. T. (2005, August 30). *Space-based power system.* Google Patents. (US Patent 6,936,760)

Roulette, J. (2022, February 11). *What is Starship? SpaceX builds its next-generation rocket.* Retrieved 13 April 2022, from https://www.nytimes.com/article/elon-musk-starship.html

Rubenchik, A., Parker, J., Beach, R., & Yamamoto, R. (2009). *Solar power beaming: From space to earth* (Tech. Rep.). Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States).

Saad, W., Bennis, M., & Chen, M. (2019). A vision of 6G wireless systems: Applications, trends, technologies, and open research problems. *IEEE Network*, *34*(3), 134–142.

Sasaki, S., Tanaka, K., & Maki, K. (2013). Microwave power transmission technologies for solar power satellites. *Proceedings of the IEEE*, *101*(6), 1438–1447.

Schlesak, J., Alden, A., & Ohno, T. (1985, December). SHARP rectenna and low altitude flight trials. In *IEEE global telecommunications conference*.

SDSC. (2016, 03 April). *Press release: Space solar power dominates the D3.* Retrieved 03 December 2021, from https://bit.ly/3EfCvM3

Shaulov, E., Jameson, S., & Socher, E. (2017). W-band energy harvesting rectenna array in 65-nm CMOS. In *Proc. 2017 IEEE MTT-S international microwave symposium (IMS)* (pp. 307–310).

Shinohara, N. (2013). Beam control technologies with a high-efficiency phased array for microwave power transmission in Japan. *Proceedings of the IEEE*, *101*(6), 1448–1463.

Shinohara, N. (2014). *Wireless power transfer via radiowaves*. John Wiley & Sons.

Shinohara, N., Hasegawa, N., Kojima, S., & Takabayashi, N. (2019). New beam forming technology for narrow beam microwave power transfer. In *Proc. 8th Asia-Pacific conference antennas propagation (APCAP)*.

Shinohara, N., & Matsumoto, H. (1998). Dependence of DC output of a rectenna array on the method of interconnection of its array elements

output of a rectenna array on the method of interconnection of its array elements. *Electrical Engineering in Japan*, *125*(1), 9–17.

Shinohara, N., Matsumoto, H., & Hashimoto, K. (2004). Phase-controlled magnetron development for sports: Space power radio transmission system. *URSI Radio Science Bulletin*, *2004*(310), 29–35.

Silva, Z. J., Valenta, C. R., & Durgin, G. D. (2020). Optically transparent antennas: A survey of transparent microwave conductor performance and applications. *IEEE Antennas and Propagation Magazine*, *63*(1), 27–39.

Skowron, J., MacMaster, G., & Brown, W. (1964). The super power CW amplitron. *Microwave Journal*, *7*.

Solaren. (2022). *About us: Our businesses.* Retrieved 11 Febraury 2022, from https://www.solarenspace.com/our-company/our-businesses/

Song, K. D., Kim, J., Kim, J. W., Park, Y., Ely, J. J., Kim, H. J., & Choi, S. H. (2019). Preliminary operational aspects of microwave-powered airship drone. *International Journal of Micro Air Vehicles*, *11*.

SpaceX. (2020, March). *Starship users guide.* Retrieved 02 April 2022, from https://www.spacex.com/media/starship_users_guide_v1.pdf

SpaceX. (2021, September). *Rideshare payload user's guide.* Retrieved 02 April 2022, from https://storage.googleapis.com/rideshare-static/Rideshare_Payload_Users_Guide.pdf

SpaceX. (2022). *Capabilities & services.* Retrieved 02 April 2022, from https://www.spacex.com/media/Capabilities&Services.pdf

Strassner, B., & Chang, K. (2003a). 5.8-GHz circularly polarized dual-rhombic-loop traveling-wave rectifying antenna for low power density wireless power transmission applications. *IEEE Transactions on Microwave Theory and Techniques*, *51*(5), 1548–1553.

Strassner, B., & Chang, K. (2003b). Highly efficient C-band circularly polarized rectifying antenna array for wireless microwave power transmission. *IEEE Transactions on Antennas and Propagation*, *51*(6), 1347–1356.

Stutzman, W. L., & Thiele, G. A. (2012). *Antenna theory and design* (3rd ed.). New York: Wiley.

Summerer, L., & Ongaro, F. (2005). Advanced space technology for 21st century energy systems: Solar power from space. In *Proc. 2005 2nd international conference on recent advances in space technologies, (RAST)* (pp. 16–23).

Summerer, L., & Purcell, O. (2009). Concepts for wireless energy transmission via laser. *Europeans Space Agency (ESA)-Advanced Concepts Team*.

Takahashi, T., Sasaki, T., Homma, Y., Mihara, S., Sasaki, K., Nakamura, S., ... Ohashi, K. (2016). Phased array system for high efficiency and high accuracy microwave power transmission. In *Proc. 2016 IEEE international symposium on phased array systems and technology (PAST)* (pp. 1–7).

Tanaka, K. (2021, 09 December). *Efforts to develop solar power satellite in Japan over the past 10 years.* Workshop: ESA's 1st International Workshop for Space Solar Power for Net Zero.

Tao, H., Su-Hui, Y., Munoz, M. A., Hai-Yang, Z., Chang-Ming, Z., Yi-Chen, Z., & Peng, X. (2014). High-power high-efficiency laser power transmission at 100 m using optimized multi-cell GaAs converter. *Chinese Physics Letters*, *31*(10), 104203.

Tesla, N. (1904). The transmission of electrical energy without wires. *Electrical World and Engineer*, *1*, 21–24.

The Economist. (2021a, 25 February). Electricity can be transmitted through the air. *The Economist*. Retrieved from https://www.economist.com/science-and-technology/2021/02/25/electricity-can-be-transmitted-through-the-air

The Economist. (2021b, 20 November). *A russian anti-satellite missile test puts the iss in peril.* Retrieved 22 November 2022, from https://www.economist.com/science-and-technology/2021/11/18/a-russian-anti-satellite-missile-test-puts-the-iss-in-peril

The Laser Institute. (2014). *Safe use of lasers* (No. ANSI Z136.1).

Turpin, T. W., & Baktur, R. (2009). Meshed patch antennas integrated on solar cells. *IEEE Antennas and Wireless Propagation Letters*, *8*, 693–696.

URSI. (2007, June). *URSI white paper on solar power satellite (SPS) systems and report of the URSI inter-commission working group on SPS.* Retrieved from https://www.ursi.org/files/ICWGReport070611.pdf

Warmann, E. C., Espinet-Gonzalez, P., Vaidya, N., Loke, S., Naqavi, A., Vinogradova, T., ... others (2020). An ultralight concentrator photovoltaic system for space solar power harvesting. *Acta Astronautica*, *170*, 443–451.

Way, S., & Lamyman, J. (2020, December). *Space based solar power as a contributor to net zero - phase 1: Engineering feasibility report* (Tech. Rep.). Frazer-Nash Consultancy.

Way, S., & Lamyman, J. (2021a, April). *Space based solar power as a contributor to net zero - phase 2: Economic feasibility* (Tech. Rep.). Frazer-Nash Consultancy.

Way, S., & Lamyman, J. (2021b, September). *Space based solar power: Derisking the pathway to net zero* (Tech. Rep.). Frazer-Nash Consultancy.

Wilson, A. R. (2019). *Advanced methods of life cycle assessment for space systems* (Unpublished doctoral dissertation). University of Strathclyde.

Xu, Y., Li, J., Tan, Q., Peters, A. L., & Yang, C. (2018). Global status of recycling waste solar panels: A review. *Waste Management*, *75*, 450–458.

Yeary, M., Palmer, R., Fulton, C., Salazar, J., & Sigmarsson, H. (2021). Update on an S-band all-digital mobile phased array radar. In *Proc. 2021 IEEE radar conference* (pp. 1–4).

York, M. C., & Fafard, S. (2017). High efficiency phototransducers based on a novel vertical epitaxial heterostructure architecture (VEHSA) with thin p/n junctions. *Journal of Physics D: Applied Physics*, *50*(17), 173003.

Zhang, B., Jiang, W., Yang, Y., Yu, C., Huang, K., & Liu, C. (2015). Experimental study on an S-band near-field microwave magnetron power transmission system on hundred-Watt level. *International Journal of Electronics*, *102*(11), 1818–1830.

Zhang, H., & Liu, C. (2016). A high-efficiency microwave rectenna array based on subarray decomposition. *Applied Science Technology*, *10*(4), 57–61.

Zhang, L.-C., & Shi, L.-X. (2022). Design of 70% PAE class-F 1.2–1.4 GHz 10 W GaN power amplifier MMIC. *Microwave and Optical Technology Letters*, *64*(4), 670–675.

Junior Management Science 8(3) (2023) 772-797

Junior Management Science

journal homepage: www.jums.academy

Green Funds and Environmental Disclosure Quality

Katharina Dormann

Humboldt-Universität zu Berlin

Abstract

I study the association between the selection of a company by a green fund and its environmental disclosure quality. Based on fund holding and environmental disclosure data of companies in the EU between 2017 and 2021 I conduct a descriptive as well as an empirical analysis. I investigate whether the environmental disclosure quality is associated with the selection by a green fund. Literature examines green funds and environmental disclosure quality separately, but the theories discussed allow for the expectation that the green fund selection and the environmental disclosure quality of companies are positively associated. I find that (i) the environmental disclosure quality of green fund investees is higher than of companies which are not selected, and (ii) the environmental disclosure quality increases further after the selection by a green fund, (iii) but this increase does not seem to be due to the selection itself but a trend of increasing environmental disclosure quality. (iv) The results suggest that green funds which rely on environmental disclosures in their selection process tend to select companies with higher environmental disclosure quality than those selected by green funds which use additional data sources besides the disclosures in their selection processes.

Keywords: Environmental disclosures; Green funds; Disclosure quality; Sustainable finance; Fund selection processes.

1. Introduction

The financial sector plays a vital role in the economic transition towards sustainability and climate neutrality (Maltais & Nykvist, 2020, 3). The European Union (EU) set out a 2030 EU Climate Target Plan that encompasses the reduction of greenhouse gas emissions to limit the consequences of climate change. The long-term goal of the EU is to reach climate neutrality by 2050 whilst growing the economy (European Commission, 2020). To reach set goals it is essential that the EU directs investments towards sustainable projects and activities. Green funds, which are mutual funds that promote environmentally conscious policies and business practices, are one instrument to direct capital towards more environmentally conscious investments (European Parliament, 2020). As of December 2020, there were more than 400 green funds who manage about 177 billion USD worldwide. In that year the global assets in green funds almost tripled. Europe is the largest market for green funds and comprises more than three fourths of the global assets in green funds (Morningstar, 2022a). If this amount of capital is directed into investments which support the transition towards sustainability and climate neutrality, they can have an immensely positive impact on reaching the EU climate targets. Especially private investors can provide large amounts of capital and can therefore support this development. That this has been recognized by the EU is mirrored in the recent implementation of mandatory education of private investors regarding sustainability in investing by investment advisors as part of the MiFID II as of 2nd of August 2022 (ESMA, 2022).

Green funds promote that they select their investments based on the environmental performance of the investment. Companies publish information on their environmental performance in corporate environmental or sustainability reports. The environmental disclosure quality is driven by the environmental performance of a company and green funds select their investment based on the environmental performance. Despite this relationship, green funds and environmental disclosure quality are usually investigated separately from each other (Lagasio & Cucari, 2019, 708). I reinforce the idea that environmental disclosures are an important instrument for fund providers to assess the business practices of companies with regard to environmental performance (Lagasio & Cucari, 2019, 701) as the environmental disclo-

DOI: https://doi.org/10.5282/jums/v8i3pp772-797

sure quality is driven by the environmental performance of a company (Gangi, D'Angelo, et al., 2016). Based on this idea I investigate the following research question. I study whether environmental disclosure quality is associated with the selection by a green funds. I assess whether companies that are selected by a green fund have a higher environmental disclosure quality when compared to those companies which are not selected.

By analysing green funds and environmental disclosure quality in the same context I aim to provide new insights into how the two subjects are interlinked. I aim to improve the understanding of the selection processes of green funds and the role of environmental disclosures in these selection processes. Overall, I expect to find that the selection by a green fund is positively associated with the environmental disclosure quality of companies, and I expect the environmental disclosure quality to further increase after a company was selected by a green fund.

For my analysis, I focus on green funds and companies within the EU as this allows for the assumption of similar legal and regulatory frameworks. To answer the outlined research question, I initially discuss the relevant terminology and legal requirements for green funds and environmental disclosure quality. Subsequently, I give an overview of prior literature on green funds as well as environmental disclosure quality and illustrate, how my research can contribute to the literature. Based on the terminology and theory I then discuss the methodology of my analysis and how I aim to investigate the potential association between green funds and environmental disclosure quality. My analysis consists of a descriptive analysis of the selection processes of green funds and an empirical analysis of the association between green funds and environmental disclosure quality. With this I aim to understand how the fund selection process and the environmental disclosure quality of firms are interlinked. After this I present my analysis and discuss my results. Based on the results I conclude by answering my research question.

2. Institutional Background

In 2020 the EU implemented a sustainable finance strategy to support the financing of the transition to a sustainable economy. This is accompanied by the implementation of the EU taxonomy, a classification system which establishes definitions for environmentally sustainable activities to, amongst other things, create more transparency for investors regarding sustainability. The framework influences both, the understanding of green funds as well as environmental disclosures in the EU as it is accompanied by the implementation of two regulations. These are the Non-Financial Reporting Directive (NFRD) which requires large companies to disclose their environmental information, and the sustainable finance disclosure regulation (SFDR) which requires investment companies to disclose how their products comply with the goals of the EU taxonomy (European Parliament, 2020).

For my analysis of the association between the selection for a green fund and the environmental disclosure quality of firms it is vital to understand how the terms green fund as well as environmental disclosure quality are defined and how they are legally regulated. In the following I therefore discuss the terminology and provide a definition for the terms which I use throughout my further analysis. Additionally, I provide insights into the current legal and regulatory requirements within the EU.

2.1. Green Funds

Green funds are mutual funds that select their investments based on environmentally conscious business practices. Green funds gained attention in the past years due to surging interests in climate change and other environmental issues (Ibikunle & Steffen, 2017, 338). There is an absence of common standards and metrics for measuring what 'green' in this context stands for and the processes based on which they select their investments differ. In general, green funds can be justified within different investment approaches (Kaufer & Steponaitis, 2021, 65). I discuss four common investment approaches for green funds in the following.

The first investment approach is SRI, which stands for socially responsible investing (Kaufer & Steponaitis, 2021, 65). The founding idea of SRI was a fight for human dignity and universal economic justice (Kaufer & Steponaitis, 2021, 78). Green funds can be defined within the SRI approach as, according to Ito, Managi, and Matsuda (2013), SRI encompasses investment vehicles which demonstrate awareness regarding social, environmental, and ethical issues. Green funds therefore focus on the environmental aspect of SRI. Alternatively green funds can be described as an investment vehicle following the ESG principles (Kaufer & Steponaitis, 2021, 78). ESG stands for environmental, social, and governance. The initial idea of ESG is attributed to the former United Nations (UN) Secretary General Kofi Annan. He urged businesses to commit to sustainable business practices (Foster, 2021, 3). Green funds can therefore also be described as mutual funds with a focus on the environmental aspect of the ESG principles. The third investment approach which justifies green funds is that of the triple bottom line. The approach comprises that business should commit to social and environmental performance in addition to their financial performance and not solely focus on profit. The concept can be broken down into three pillars, being people, planet, and profit (Elkington, 1998, 22).Within this concept green funds can be described as investment vehicles with a focus on the planet-pillar of the approach. Lastly, green funds can be seen as an investment vehicle following the impact investing approach (Kaufer & Steponaitis, 2021, 78). This strategy seeks financial returns whilst creating a positive environmental or social impact (Clarkin & L. Cangioni, 2016, 137-138). Within this, green funds can be described as mutual funds with a focus on a positive environmental impact.

The terms SRI, ESG, triple bottom line, and impact investing are often used interchangeably but they have different origins and practices (see Table 1). SRI involves the selection of investments based on sustainability criteria. ESG considers

the environmental, social, and governance aspects alongside traditional financial measures. The triple bottom line formulates the three pillars, people, planet, and profit as minimum requirements. Impact investing aims to help businesses to achieve a positive impact. A shortcoming of all approaches is the lack of consistent terminology as well as uniform measurement standards (Kaufer & Steponaitis, 2021, 77). Whilst green funds can be justified in all of these different investment approaches, I conclude that green funds can be defined as mutual funds that select their investments based on the environmental performance of a firm alongside traditional financial or performance indicators.

As illustrated above, green funds follow different investment approaches. Therefore, it is plausible that green funds are dissimilar in their characteristics. Inderst, Kaminker, and Stewart (2014) illustrate that the funds differ regarding their dimensions (number of stocks, average size, liquidity, or sector breakdowns) and their selection criteria. Green funds can have a sectoral or thematic focus (e.g., alternative energy, clean technology, or carbon emissions) or can also not focus on specific aspects of green investment. With this understanding of green funds in mind, I describe the legal requirements for green funds in the EU in the following.

2.1.1. Legal Requirements for Green Funds in the EU

The EU states that to reach the climate targets for 2030 and to direct investments towards sustainable projects and activities such as green funds, a clear definition of what can be called green or sustainable is needed. Therefore, the EU started to implement the EU taxonomy in 2020 (European Parliament, 2020). This taxonomy implements requirements for sustainable finance and therefore also green funds as well as further legal requirements which can be guidelines to understand what can be classified as 'green'. Therefore, I discuss the regulations which are part of the taxonomy and relevant for green funds in the following.

In 2019 the European Parliament passed the Sustainable Finance Disclosures Regulation (SFDR). Since March 2021 the providers of financial products have to comply with this regulation which comprises technical standards to be used by financial market participants when disclosing sustainability related information (European Parliament, 2019). Investment products according to the definition in Article 2 of the SFDR encompass investment funds. Therefore, the requirements apply to providers of investment funds and are relevant for green funds. Articles 6, 8 and 9 of the SFDR classify financial products into three different investment strategies (see Table 2). Article 6 of the SFDR covers products that do not integrate any sustainability criteria. They are allowed to be sold in the EU but are clearly labeled as non-sustainable. Green funds therefore do not fall under the regulations of that article. Articles 8 and 9 cover products which promote environmental or sustainable investments (European Parliament, 2019).

Article 8 of the SFDR comprises funds that promote environmental and social characteristics. Products that are labeled as compliant with Article 8 promote financial products which are selected based on environmental or social criteria, or a combination of both and additionally have to ensure that their investments follow good governance practices (European Parliament, 2019).

Article 9 comprises investment funds that include products that target specific sustainable investments and applies when a product has a sustainability target as its objective. Examples of such objectives are products that target green investments, the reduction of carbon emissions or the achievement of the climate goals of the Paris Agreement. Furthermore, Article 9-funds have to incorporate criteria of good governance in their investment strategy (European Parliament, 2019).

Products that want to comply with either Article 8 or 9 are also required to assess the fund portfolio against the principle of "do no significant harm" by considering the principal adverse sustainability impact indicators (PASIs). The fund providers have to ensure that the products, their fund invests in, do not cause negative impacts on for example the environment or human rights. How the PASIs are assessed, is up to the fund providers. The fund providers have to incorporate considerations regarding minimum social safeguards of their investments which are specified in the EU taxonomy (European Parliament, 2019, 2020).

If a fund is labeled as a product complying with Article 8 or 9, the fund providers have to disclose information regarding how the financial products included in the fund comply with the legal requirements. As of now there are no detailed specifications regarding the disclosure format. In general, the funds have to provide information on how they meet the characteristics required by the SFDR and additionally, if they measure their targets by comparison to an index, information on how that index is compliant with the characteristics. The SFDR does comprise reporting templates for the disclosures regarding compliance with Articles 8 and 9, the mandatory use of these templates however will start to apply on the 1st of January 2023 with disclosure regarding PASIs at entity level being further delayed until 30th of June 2023 (European Parliament, 2019). Based on the legal requirements I summarise that green funds in the EU are investment funds which comply with either Article 8 or 9 of the SFDR and moreover set their focus on environmental criteria in the selection of their investments.

Despite the SFDR together with the EU taxonomy providing some guidelines, the regulations do not yet provide a uniform definition on what is classed as an environmental or sustainable target. The regulations are leaving room for interpretation on which targets can be interpreted as green or sustainable and how these objectives need to be measured and benchmarked. Therefore, the asset allocation strategies and selection processes of these funds differ. In the following I illustrate the different investment approaches and selection processes of green funds.

2.1.2. Selection Approaches of Green Funds

In their selection processes funds use different metrics to assess environmental performance. Some providers select

Table 1: Investment Approaches for Green Funds

Source: Derived from Kaufer & Steponaitis, 2021, 65-78. Notes: This table provides an overview regarding the different investment concepts and approaches within which the concept of green funds can be justified.

Investment Concept	Key Ideas and Practices
SRI	Connect investments to socially responsible values and positive change of corporate practices.
ESG	Membership group with public commitment to sustainable business practices.
Triple Bottom Line	Reporting practices to create transparency for the social and environmental impact of companies.
Impact Investing	Align investment opportunities with impact objectives for a positive social change.

Table 2: Overview of SFDR Articles for Funds

Source: Derived from SFDR Articles 6, 8 and 9.
Notes: This table provides an overview regarding the level of integration of green objectives required by the SFDR for investment funds. The articles are sorted from lowest to highest level of green integration. Additionally, I show the names of the classes as which the fund are described according to the SFDR.

Article	Requirements	Class
6	None. Includes all managed funds.	None
8	Promotion of environmental or social characteristics.	Light Green
9	Sustainable investment objective.	Dark Green

investments based on qualitative metrics such as the operation in a green sector (e.g., sustainable energy), others use quantitative measures and invest in the firms with for example the largest contribution to reducing emissions within one industry. Consequently, it is not surprising, that the actual funds are very different in their characteristics (Inderst et al., 2014, 26-28). Whilst some funds implement an environmental research team tasked with the identification of firms that match the environmental criteria, others only focus on the exclusion of firms that fulfil negative criteria (Stuart & Bioy, 2021, 4). In general, firms that are selected for a green fund are required to fulfil both, financial and environmental criteria. This aligns with the ideas of the investment approaches which I present before in Table 1.

Overall, five main categories of green fund strategies can be identified (see Table 3). These are low carbon, climate conscious, climate solutions, clean energy/tech, and green bond (Stuart & Bioy, 2021, 4). Low carbon funds invest in firms whose carbon intensity or carbon footprint is lower when compared to a benchmark index. Climate conscious funds select firms that consider the challenges of climate change in their business strategy and therefore, either align with a transition towards a low-carbon economy or provide carbon solutions. Climate solution funds focus on firms whose products provide solutions for the challenges of climate change. Clean energy/tech funds invest mostly in green energy solutions such as renewable energies or smart power management technologies (Stuart & Bioy, 2021, 4-5). Green bond funds invest in debt instruments with positive environmental and or climate benefits.

In difference to the other four fund categories outlined before, the International Capital Market Association (ICMA) provides detailed requirements for eligible projects for green bonds which are called the green bond principles (GBP). Large fund providers are voluntarily members of this association and therefore required to apply the green bond principles (ICMA, 2021). The other green funds categories do not have voluntary or legal restrictions.

Despite the different fund categories in which green funds can roughly be classified, there are no uniform definitions of green funds. The main criteria for a green fund are that it selects investments that support the transition into sustainable and climate-neutral economy and support firms or projects with positive environmental or climate benefits.

2.2. Environmental Disclosures

The environmental performance of a company becomes more and more important for stakeholders and shareholders. In 2020 a 77 % majority of publicly listed companies in Europe has adopted sustainability reporting (KPMG, 2022, 10). Similar to the term green funds, the term environmental disclosures does not have a standard and uniform definition. Environmental disclosures can be explained as a sub-group of disclosures within different normative frameworks (Hahn & Kühnen, 2013, 7). On the one hand, environmental disclosures can be described as a part of sustainability disclosures. Sustainability disclosures contain three dimensions, economic, environmental, and social disclosures (Lozano & Huisingh, 2011, 103). On the other hand, environmental disclosures can be understood as a part of corporate social responsibility (CSR) reporting. CSR is defined as the responsibility of an organization for the impact of its decisions on society and the environment (European Commission, 2011).

Following these definitions, environmental disclosures should comprise information on the company's impact on the environment. This means that the environmental disclosures

Table 3: Categories of Green Funds

Source: Derived from Stuart & Bioy, 2021, 4-5 and ICMA, 2021.
Notes: This table provides an overview regarding the main categories of green funds.

Category	Investment Concept
Low Carbon	Firms with reduced carbon intensity/footprint relative to benchmark index.
Climate Conscious	Firms that consider climate change in their business concept or provide carbon solutions.
Climate Solutions	Firms that provide climate change solutions.
Clean Energy/Tech	Firms that contribute to the energy transition.
Green Bonds	Debt instruments that finance green projects.

can comprise information on for example carbon emissions, carbon footprint, waste production, or the energy and water consumption (Braam, de Weerd, Hauck, & Huijbregts, 2016, 724). The level and nature of the environmental information disclosed by the firms varies (Hahn & Kühnen, 2013, 6). To understand what environmental disclosures need to comprise and how environmental disclosures should be organized, I describe the legal requirements for environmental disclosures in the EU in the following section.

2.2.1. Legal Requirements for Environmental Disclosures in the EU

In the last years there was a steep increase in companies publishing environmental disclosures in the EU. The EU undertook several steps to standardize the requirements for sustainability reporting of which environmental disclosures are a part. In 2014 the EU published the Non-Financial Reporting Directive (NFRD) which, in short, requires all publicly listed companies and financial institutions with more than 500 employees to report non-financial information. The NFRD was implemented in national law by the EU member states between 2015 and 2018. The NFRD requires these companies to publish information on environmental and social matters as well as treatment of employees, respect for human rights, anti-corruption, bribery, and the diversity of company boards. Additionally, the companies are required to publish information on the due diligence processes within the corporation. With regards to environmental disclosures the NFRD requires information on the current and foreseeable impacts of the business on the environment, health and safety, the use of energy, greenhouse gas emissions, use of water and air pollution. This is only enforced so far, that statutory auditors are required to assure that the non-financial information has been provided. Member states are allowed to require an independent assurance of the information via their national law (European Parliament, 2014). In 2017 the European Commission published additional guidelines to support companies in the disclosure of environmental and social information, followed by a guideline regarding the reporting of climate-related information, but the application of these guidelines is not mandatory (European Commission, 2019). The NFRD itself does not provide a system of reporting standards and the implementation into national law by the EU member states varies (European Commission, 2017, 2019).

While the current legal requirements in the EU aim to increase the relevance, consistency and comparability of information disclosed and provide orientation, they do not yet provide a system of detailed standards for the disclosure of sustainability information, which would be an important step towards more comparability of sustainability disclosures in the EU (EFRAG, 2022). The EU realized the need for a further standardization of environmental and sustainability disclosures. Other regulatory initiatives are currently under development. One is for example the proposal for a Corporate Sustainability Reporting Directive (CSRD), which would amend existing reporting requirements of the NFRD and aims for an adoption of EU sustainability reporting standards which are already being developed by the European Financial Reporting Advisory Group (EFRAG) (European Parliament, 2021). Additionally, an EU-wide system for the classification of sustainable activities was published in 2020 as part of the EU taxonomy and is being implemented in several steps (European Parliament, 2020). Whilst these planned regulations are not yet implemented or only partly implemented, they already increased the momentum for sustainability-related reporting in the EU (KPMG, 2022, 12).

2.2.2. Requirements for Disclosure Quality

The current legal requirements do provide general statements regarding what environmental disclosures should contain but they do not provide information regarding the level of detail of disclsoures and which quality criteria the environmental disclosures need to fulfil. Besides the legal requirements which were outlined before, firms in the EU are allowed to adopt voluntary sustainability reporting standards as long as these standards also fulfil the legal requirements (European Parliament, 2014). Examples of voluntary standards with a widespread adoption in the EU are the reporting principles of the Global Reporting Initiative (GRI) and the recommendations of the Taskforce for Climate-Related Financial Disclosures (TCFD) (KPMG, 2022, 5-6).

The GRI develops voluntary standards for sustainability reporting which provide more detailed reporting standards than the current legal requirements in the EU. The GRI standards are the dominant voluntary reporting standards worldwide (KPMG, 2022, 25). The standards of the GRI also allow for an understanding of what quality of environmental disclosures can be defined as. The GRI defines accuracy, balance,

clarity, comparability, completeness, timeliness, and verifiability as principles for the quality of disclosures and provides detailed standards on how firms should disclose environmental information in order to meet the quality principles (GRI, 2022). The TCFD was launched by the Financial Stability Board and provides recommendations on how firms should disclose climate-related financial information in order to enable informed and efficient investment decisions. The TCFD recommendations are applied by more than one in five firms in Europe (KPMG, 2022, 39). Similar to the GRI guidelines the TCFD also encompasses principles for disclosure quality, being relevancy, completeness, clarity, consistency, comparability, reliability and timeliness (TCFD, 2017, 2018). Other voluntary standards for environmental disclosures such as those of the International Sustainability Standards Board (ISSB) or the Sustainability Accounting Standards Board (SASB) use similar terms as the GRI or TCFD when describing their principles for environmental disclosure quality (ISSB, 2022; SASB, 2020).

Voluntary environmental disclosure initiatives have a similar understanding on what criteria environmental disclosures need to fulfil to be of quality. In the EU there are no standards for environmental disclosures implemented yet, but the EFRAG published working papers for European Sustainability Reporting Standards (ESRS) in March 2022 which cover environmental disclosures. The working paper of the ESRS also covers characteristics of disclosure quality. These are defined as relevance, faithful representation, comparability, verifiability, and understandability (EFRAG, 2022). Despite the differences in terminology, the voluntary standards and the drafts for future EU legislation show a similar understanding of which characteristics environmental disclosures need to fulfil to be described as disclosures of quality, for an overview on the terminology used by the voluntary standards providers as well as the EFRAG see Appendix I. Overall, I summarise that characteristics for the quality of environmental disclosures are not only the extent of the disclosed information but also their credibility plays an important role.

3. Literature Review

To get an understanding of what prior literature discusses regarding green funds and environmental disclosure quality I in the following provide a literature review on these subjects. I aim to reveal trends, relations, and potential gaps in the literature and evaluate, how my research can contribute to literature. I start with a review of literature on green funds followed by a review of literature on environmental disclosure quality. As I outline before, both terms, green funds and environmental disclosure quality, do not have a uniform definition but a wide variety of descriptions and associated terms. Thus, I believe it is useful to initially review the literature regarding my research separately out of the perspective of green funds and out of the perspective of environmental disclosure quality before I outline the interlinkage of the green funds and the environmental disclosure quality. Lastly, I discuss prior literature regarding a potential association between green funds and environmental disclosure quality. Based on my results in this review I formulate my expectations for the empirical analysis.

3.1. Green Funds

Prior literature indicates an increasing relevance of environmental and climate issues for capital markets. Prior research illustrates that institutional investors, such as fund providers, play a major role in encouraging a transformation into an environmentally friendly economy. Therefore, green funds are one tool to redirect investments into companies and projects which support a transition towards a more sustainable economy and to focus on sustainable long-term firm value rather than short-term profit maximization (Busch, Bauer, & Orlitzky, 2016, 310).

From a theoretical perspective, prior literature distinguishes an economic and an ecological perspective on green investments, which covers green funds (Busch et al., 2016, 308-309). From an economic perspective, profits need to be accumulated on the basis of long-term strategies and need to be responsibly related to the real term increase of economic value in order to be sustainable or green. Also, it requires that profits are not based on corruption and that elementary needs are not threatened. From an ecological perspective, the profit-making of green investments needs to be consistent with increasing resource productivity and usage of renewable resources, recycling and reuse of materials as well as the preservation of global and regional ecosystems. Therefore, proposed investments of green funds must fulfil both, financial and environmental criteria (e.g., Busch et al., 2016, Hoffmann, Scherhorn, & Busch, 2004, Ryall & Riley, 1996). This implies that a firm that is deemed as green would not be selected by a green fund if it does not fulfil the criteria of financial performance (Ryall & Riley, 1996, 234).

How green funds are supposed to define environmental criteria for the assessment of green behaviour and how to select their investments is, even though it is not discussed by many papers, controversially discussed in prior literature (Arribas, Espinós-Vañó, García García, & Oliver-Muncharaz, 2019, 1642). One of the main challenges when assessing whether a company is green and should be included in a green fund is a lack of consensus on the exact meaning of being ‘green’. Also, this concept is often confused or mixed with other concepts in the universe of sustainable finance. And even if the same definition is applied, fund providers and other players such as rating agencies still apply different metrics for the measurement of the criteria (Capelle-Blancard & Monjon, 2012, 244).

In literature, two main approaches for the selection of green investments are discussed. These are negative and positive screening. Negative screening means that exclusion criteria are used to assess whether an investment is deemed as green (e.g., exclusion of coal or oil-based power generation). The environmental performance itself is not analysed, just the economic situation. Minimum standards are defined

and if the company does not comply with these it is excluded (Arribas, Espinós-Vañó, García García, & Oliver-Muncharaz, 2019, 1644). The negative screening approach is criticized in the literature as the exclusion criteria are not able to cover all non-green practices and activities which leads to companies, which clearly undertake non-green activities, being included in green funds whilst companies which actually are green are not selected if simple negative screening criteria are used (e.g., Arribas, Espinós-Vañó, García García, & Tamosiuniene, 2019, Hellsten & Mallin, 2006). Positive screening is an approach where investment managers implement systematic environmental factors in their financial analysis and the investment selection process. They assess the environmental performance of the companies and usually form a score based on this. They select those firms with the highest score for their portfolio (Arribas, Espinós-Vañó, García García, & Oliver-Muncharaz, 2019, 1644). This approach is discussed in literature due to its complexity. Different variables must be assessed and measured before they can be weighted and combined into a score. This leads to points of criticism such as a lack of standardization and credibility of such scores and the input data as well as a lack of transparency and distortions due to a potential subjective bias of the score designers (Windolph, 2011, 42-49).

Overall, the literature on green funds is limited as most studies focus on various components of sustainable investments and blend different terms under the umbrella of sustainability (e.g., Ibikunle & Steffen, 2017, Ito et al., 2013). Few contributions cover intra-industry, intra-country or green versus conventional fund performance analyses (Lagasio & Cucari, 2019, 708). But the literature highlights the challenges regarding green funds, which are the lack of a uniform definition of what is green and the problems in the assessment of this for potential investments.

3.2. Environmental Disclosure Quality

The literature on environmental disclosures extensively discusses environmental disclosures and their quality in the context of various theories and setups. In the following I aim to provide an overview of the relevant literature and its implications for my analysis of the interlinkage between environmental disclosure quality and green funds.

In prior literature environmental performance is described as a driver for environmental disclosure quality. The companies disclose their environmental performance in order to fulfil stakeholder claims (Gangi et al., 2016, 1399). Prior literature on environmental performance and environmental disclosure quality identifies different incentives for companies to present environmental disclosures of quality. These incentives are based on forces relating to the legitimacy of the firm and institutional-oriented forces (Maltais & Nykvist, 2020, 6).

Frequently mentioned theories used to explain environmental disclosure quality are institutional theory, stakeholder theory, legitimacy theory and signaling theory (e.g., Maltais & Nykvist, 2020, Braam et al., 2016, Hahn & Kühnen, 2013). Following Fernando and Lawrence (2014) these theoretical approaches can be described as a set of theories that predict similar or complementary incentives regarding the environmental disclosure quality of firms. Due to similar theoretical predictions, it is difficult to sharply differentiate between the different theories in the context of environmental disclosure quality (Maltais & Nykvist, 2020, 6-7). In short, the theories predictions promote the idea that firms are incentivised to comply with societal norms and values and use their environmental disclosures to comply with values and norms regarding their environmental performance (Campbell, Craven, & Shrives, 2003, 559).

Following from this idea, companies' efforts to provide environmental disclosures of high quality can be explained by the company being incentivised to secure their legitimacy by operating within societal norms (Maltais & Nykvist, 2020, 7). A part of these norms and values is that companies are expected to operate on a high level of environmental performance. Additionally, the companies want to demonstrate accountability for their business practices by providing information within their environmental disclosures. Also, in a situation of asymmetric distribution of information, which is present in the relationship of a company with its stakeholders and shareholders, companies voluntarily disclose information to differentiate themselves from peers or competitors (Hahn & Kühnen, 2013, 14). To achieve this, they provide credible information which cannot be replicated by inferior environmental performers which leads to the expectation of literature that superior environmental performers also provide environmental disclosures of higher quality when compared to inferior environmental performers (Braam et al., 2016, 725). Therefore, based on these theories, environmental performance can be described as a driver of environmental disclosure quality (Gangi et al., 2016, 1399). This is mirrored in the phenomenon observed by literature that companies do not only disclose environmental information driven by regulatory demands but also provide voluntary environmental disclosures to convey their compliance with the societal norm of environmentally friendly business practices (Pérez-López, Moreno-Romero, & Barkemeyer, 2015, 722).

Whilst the theories discussed in prior literature offer arguments for why superior environmental performers provide environmental disclosures of higher quality when compared to inferior performers, prior literature does not neglect the problems coming with this argumentation. Within the set of theories companies with an inferior sustainable performance are expected to voluntarily disclose more environmental information to distract from their inferior performance (Braam et al., 2016, 726). Overall, results from prior literature imply that environmental disclosures are primarily used by firms to improve the environmental image and reputation of the firm as they found a positive impact of disclosing information regarding positive environmental performance internally as well as externally for the firms (e.g., Birkey, Michelon, Patten, & Sankara, 2016, Michelon, Pilonato, & Ricceri, 2015). But to gain this reputational benefit the environmental disclosures need to be credible, and the information provided needs to be matched by the actions of the firm. Therefore,

prior literature identifies credibility as the main reason for the positive reputational benefits of environmental disclosures by firms (Birkey et al., 2016, 144). Accordingly, prior literature claims that if a company acquires external assurance of the information presented in its environmental disclosures, this enhances the credibility of the disclosures and therefore allows for positive reputational benefits for the firm (e.g., Del Giudice & Rigamonti, 2020, Birkey et al., 2016, Chen, Srinidhi, Tsang, & Yu, 2016, Kolk & Perego, 2010). This implies that an increasing extent of disclosures does not necessarily enhance the environmental disclosure quality, but the credibility of the information does (Fernandez-Feijoo, Romero, & Ruiz, 2014, 54).

The importance of the credibility of the disclosed information on environmental performance by companies is mirrored in the approaches to measure environmental disclosure quality which are provided in prior literature. These framework approaches usually aim to measure environmental disclosure quality in a score and the disclosure items which are used in these frameworks can be assigned to two categories: hard disclosures (e.g., quantitative environmental performance indicators), and soft disclosures (e.g., vision and strategy claims). Hard disclosures are items that are verifiable and difficult to mimic whilst soft disclosures are of a general nature and difficult to verify (e.g., Michelon et al., 2015, Hahn & Kühnen, 2013, Clarkson, Overell, & Chapple, 2011). In general, the frameworks cover the adoption, extent, and the credibility of the environmental disclosures as central components for environmental disclosure quality.

Even though the frameworks to assess environmental disclosure quality which can be found in prior literature use similar disclosure items to assess disclosure quality, the usefulness of these frameworks itself is controversially discussed in prior literature. As regulations regarding environmental disclosures on an international level are still in their infancy, the environmental information disclosed by companies varies. This can impact the reliability of environmental disclosure scores as the frameworks, which assess the scores, have no uniform definition of environmental disclosure quality and which criteria need to be fulfilled within the score (Del Giudice & Rigamonti, 2020, 5672). Additional controversial issues are the complexity regarding the terminology and definition of what is green, the criteria of choice used in the frameworks, or the judgement of whether a criterion was met (e.g., Diez-Cañamero, Bishara, Otegi-Olaso, Minguez, & Fernández, 2020, Semenova & Hassel, 2015, Chatterji, Levine, & Toffel, 2009). Especially measurement divergence regarding disclosure items seems to account for a large part of discrepancies in environmental disclosure scores. Also, environmental disclosure scores might adopt different definitions of environmental performance and its determinants (Del Giudice & Rigamonti, 2020, 5673).

3.3. Literature Gap: Green Funds and Environmental Disclosure Quality

Prior literature for the most part does not focus on the environmental disclosure quality in the context of green funds but analyses both separately (Lagasio & Cucari, 2019, 708). But environmental disclosures are discussed as a key factor based on which the green funds select their investments. Despite environmental disclosures being mentioned as one problem in the assessment of a firm's environmental performance due to problems such as greenwashing, which is the practice of disclosing misleading information regarding the company's environmental commitment, environmental disclosures of a firm are described as the fundament on which the fund's examination of a company's environmental performance is based. Evidence in prior literature shows that despite the differences in the selection processes, green funds always initialize their selection process based on the environmental information disclosed by the firm itself (Ryall & Riley, 1996, 236-238).

Based on the theories which are discussed in prior literature it becomes visible that there is potential for an association between green funds and the environmental disclosure quality of companies. I argue, following the theories discussed in prior literature, that companies which are selected for a green fund have a higher environmental disclosure quality when compared to companies which are not selected. If they would not credibly signal a high level of environmental performance, this would undermine the legitimacy of their inclusion in a green fund and after their selection they need to continue to signal the legitimacy of their selection by a green fund to stay selected. Green funds select their investments based on the environmental performance of the companies, which is presented within the environmental disclosures. Therefore, the quality of the environmental information disclosed by the companies is relevant in their selection process. At the same time it is desirable for companies to be selected for a fund. If a company's stocks or bonds are selected for a fund this is a chance for the firm to gain reputational and financial benefits (Bancel & Mittoo, 2009, 846). Additionally, the selection by a green fund, is a chance to attract investors who want to benefit from both, the instant diversification in mutual funds and the chance to invest in sustainable companies or projects (Bassen, Gödker, Lüdeke-Freund, & Oll, 2019, 63).

Therefore, I expect to find an association between the selection for a green fund and a higher environmental disclosure quality of selected companies when compared to companies that were not selected. Furthermore, I expect the environmental disclosure quality of a company to increase after the selection by a green fund. To illustrate, how I approach my analysis of the association between green funds and environmental disclosure quality, I discuss my methodology in the following.

4. Methodology

In the following I describe how I address my research question. Prior to my analysis I gather data samples to construct a viable timeframe. I decide to conduct my analysis for the years of 2017 to 2021, therefore I cover two years before and after the implementation of the SFDR in the EU. Due to

the implementation of the SFDR regulation in the EU many funds were newly launched or restructured in order to comply with either Article 8 or 9 of the SFDR and to be classified as green. This allows me to identify companies which are initially selected for a green fund or deselected. Such observations help me to identify a potential association between the green fund selection and the environmental disclosure quality of firms. Based on this understanding I conduct my data collection which is the fundament for my analysis. I illustrate my data collection process as well as the steps of my analysis in the following.

4.1. Data Collection

To address my research question, I require data for the holdings of green funds as well as for the environmental disclosure quality of the companies which are investees of the green funds. Besides that, I have to decide on conditions based on which I construct my sample as well as which key assumptions I want to make. Therefore, I illustrate the data collection process for the green fund data as well as the environmental disclosure quality data in the following.

4.1.1. Green Fund Data

For my analysis of green funds and environmental disclosures in the EU I need to identify a sample of green funds and collect data on their holdings over the selected timeframe. I focus on funds which are classified according to Article 8 or 9 of the SFDR and have an additional focus on environmental or climate issues or at least do not explicitly exclude environmental or climate objectives in their asset allocation strategy. Additionally, I choose the green funds with the largest inflows in the fourth quarter of 2022 in the EU, effective date 31st of December 2021. These funds have net flows of more than 400 million EUR which is significantly higher than the inflows of other European green funds. High inflows highlight that investment activities took place in these green funds during this period which makes these funds useful examples when looking at the current state of selection processes of green funds. Besides these the funds have to fulfil the following criteria to be part of my analysis. I include open-ended mutual funds, primary funds and equity funds which are active and have their domicile in the EU. I do not include bonds which are issued for a specific project as they do not represent a company. I choose to focus on the EU as this allows for the assumption of a similar regulatory and legal framework. For an overview of the resulting funds see Table 4.

I choose to take the 31st of December 2021 as the effective date for the fund inflows, as the fund providers usually publish their annual fund reports with an effective date of 31st of December. Therefore, I have the same effective date for the fund inflow data, based on which I select my sample of green funds, as well as for the fund holding data. Additionally, this means that I use the most recent data for fund holdings which is currently publicly available. I manually collect the data of the funds' holdings out of their annual fund reports to ensure that my data set is as complete as possible and covers the funds I selected. These fund reports are provided yearly and show all holdings per fund as of the effective date. Moreover, this manual data collection allows for me to simultaneously collect further data points such as the funds' investment approaches and objectives and information on their selection processes.

To notice, the fund providers which hold the funds I selected as examples for my analysis are all under the top 20 of asset managers which provide Article 8 or 9 SFDR fund assets as of 31st of December 2021 (Morningstar, 2022b, 17). This allows for the assumption that the selection processes and asset allocation strategies of the selected funds are relevant examples for the selection processes of green funds in the EU and not specific to a minority of fund providers.

As a control group, I use a set of comparable firms which are not part of a green fund. I select the control group out of conventional mutual funds and again start my selection based on the top ten funds based on inflows as of the fourth quarter of 2021. Then I eliminate holdings which are selected for a green fund in my sample. Based on the resulting company universe I construct a sample control group of comparable size and fundamental financials when compared to my treatment group to prevent as best as possible that my empirical results are driven by other determinants than the selection for a green fund. This results in a balanced panel data sample.

4.1.2. Environmental Disclosure Quality Data

To approach the quality of the environmental disclosures I build on frameworks of environmental disclosure quality which can be found in prior literature (e.g., Braam et al., 2016, Michelon et al., 2015, Clarkson, Li, Richardson, & Vasvari, 2008). These framework cover indicators for the adoption, extent, and credibility of the environmental disclosures and combine them into one environmental disclosure score (EDS) which is in line with the requirements of disclosure quality as discussed in Chapter 2.1.2 and presented in prior literature as illustrated in Chapter 3.2. I use this score as a proxy for the environmental disclosure quality. The disclosure items which I assess using the framework cover hard and soft disclosures. Within the score, hard disclosure items are considered in an overweight position. The framework and the used indicators for my score are presented in Appendix II. I base my framework on the frameworks in prior literature. As the most recent framework is from 2016, I review and update the indicators based on the current requirements for environmental disclosures from voluntary disclosure standards and the standards drafts by the EFRAG for the EU regulations. The sources for this are indicated in Appendix II as well.

To determine the score, I use data from the Thomson Reuters EIKON ESG database. This database covers the indicators which are part of my framework for the environmental disclosure score. This data is the main source for the environmental disclosure score data to avoid a subjective bias. I only fill in missing data manually if necessary for the analysis. Based on this data I determine the environmental disclosure scores of the companies in my sample over the timeframe

Table 4: Selected Green Funds

Source: Derived from Morningstar, 2022a, 2022b, data as of 31st of December 2021.
Notes: This table reports the funds which I choose for my descriptive and empirical analysis based on the criteria outlined above as well as the respective fund providers. It additionally indicates whether the fund is classified as Article 8 or 9 of the SFDR.

	Fund Name	Fund Provider	SFDR Article
1	SRI Euro Quality	DNCA	8
2	ESG Multi-Asset Fund	BlackRock	8
3	Global Impact	UBS	8
4	Global SDG Engagement Equities	RobecoSAM	9
5	Sustainable Energy Fund	BlackRock	9
6	Sustainable Global Thematic Portfolio	AllianceBernstein	9
7	Global Climate and Environment Fund	Nordea	9
8	Worldwide Positive Change Fund	Baillie Gifford	9
9	Global Sustainable Equity Fund	Mirova	9

from 2017 to 2021 which together with the green fund data results in a balanced panel.

4.2. Data Analysis Process

I start my analysis with a descriptive analysis of the green fund selection processes. For this, I analyse the information of the fund providers and managers on the selection process as published for the selected green funds. All sources used for this descriptive analysis are presented under 'Additional Resources' which follows my reference list. I aim to provide insights into the selection processes and their differences as well as the interlinkage of this process with the environmental disclosure quality of the investees. Furthermore, I aim to create categories of green fund selection processes in order to categorize them based on the level of detail of their assessment of the environmental performance of the investees in the selection process. I expect such a categorization to be interesting to include in my empirical analysis. With the understanding of this process and its interlinkage to environmental disclosure quality, I then conduct an empirical analysis in order to answer my research question.

My empirical analysis consists of several regression models. To initialize my analysis, I implement a regression model with the environmental disclosure score as the dependent and the selection for a green fund as the independent variable. I control for company and year fixed effects. In that setting, I analyse whether my expectation that the selection by a green fund is associated with environmental disclosure quality can overall be confirmed and whether the companies which are selected for a green fund have a higher environmental disclosure quality when compared to those companies which were not selected. Furthermore, I want to detect whether the environmental disclosure score increases after the selection by a green fund and is not just generally higher when compared to those firms which were not selected. Based on these two main regression models I additionally implement regression models in which I use the components of the environmental disclosure score (adoption, extent and credibility) as separate dependent variables as well as models where I use the green fund types which I identify in my descriptive analysis as independent variables. Additionally, I use firm and year fixed effects as well as both together to control for unobservables. When I use both, the standard errors are clustered at firm level. In the models I also use determinants of environmental disclosure quality as control variables. Based on prior literature I use indicators for company size, profitability as well as leverage as their influence on the environmental disclosure quality is widely acknowledged in literature (e.g., Lagasio & Cucari, 2019; Braam et al., 2016; Hahn & Kühnen, 2013; Clarkson et al., 2011).

5. Analysis Results

In the following I present the results of my analysis. I start by illustrating my findings from the investigation of the green fund selection processes. After this I describe the results of my empirical analysis. This is followed by a depiction of the development of the relationship between green funds and environmental disclosures over time. Lastly, I provide a discussion of my results which covers the interpretation of my results and findings as well as potential limitations of my analysis.

5.1. Descriptive Analysis of Green Funds and Environmental Disclosure Quality

As described before, funds which are labeled as green in the EU fall under either Article 8 or 9 of the SFDR. In order to fulfil the requirements of the regulations, the funds have to screen and assess potential investments. To decide in which products they want to invest with their green funds, fund providers have to consider the environmental performance of the potential investments. There are no specific legal requirements regarding the process of how green funds have to select their investments. Therefore, the selection processes as well as asset allocation strategies of green funds differ. Whilst some funds implement an environmental research team tasked with the identification of companies which match their environmental criteria, others only focus

on the exclusion of firms that fulfil negative criteria (Stuart & Bioy, 2021, 4).

To illustrate how the selection processes of green funds differ and how this is interlinked with the environmental disclosure quality of companies in which green funds invest, I investigate the selection processes of the green funds in my sample as well as the results of these processes the following, for the selected green funds see Table 4.

5.1.1. Selection Processes of Green Funds

At first, I take a look at the investment approaches and fund categories which the selected green funds can be categorized into to look for potential patterns regarding the approaches and categories. For definitions of the investment approaches see Table 1 and for a description of the green fund categories see Table 3.

All of the considered funds follow specific investment approaches. In the asset allocation strategies and key investor information documents (KIID) I find SRI, ESG and impact investing as investment approaches which are applied by the funds. Most common is the SRI investment approach which is applied by four of the nine funds (see Table 5).

The green funds all can be assigned to a corresponding green fund category. The most common category, concerning six of the nine funds, is climate conscious, which can also be described as the most general category of green funds as it does not focus on for example a specific industry. The other categories are clean energy/tech and climate solutions. With regards to the green fund categories I notice that only funds which are classified according to Article 9 SFDR, which is also described as dark green (see Table 2.2), belong to a green fund category different than climate conscious. Therefore, in my sample, only Article 9 SFDR green funds have a more specific investment scope than just incorporating general environmental or climate criteria.

Based on these differences in investment approaches and fund categories I expect to find a pattern regarding how elaborate the selection processes of the corresponding funds are. Therefore, I take a closer look at the separate steps the funds conduct in their selection processes in the following, for an overview see Table 6.

All of the funds perform negative screening, usually as the first step in their selection process. As discussed in the literature review, negative screening describes a strategy by which securities are excluded as potential investments when they are not aligned with the values of the fund's strategy. The green funds usually exclude business such as coal mining, fossil-fired power generation, conventional oil and gas, production of oil sands, arctic drilling, or production of palm oil. The only exception here is the fund Sustainable Global Thematic Portfolio by Alliance Bernstein, which does conduct negative screening, but only for controversial weapons, and not for any climate or environmental criteria.

Interesting to notice here is also that several funds explicitly exclude companies that produce gas or nuclear power. In a recent decision from the 6th of July 2022 the European Parliament voted to classify gas and nuclear energy as green within the EU sustainability taxonomy after long and controversial debates. This could imply that retail investors either expected the European Parliament to take a different decision or they believe that gas and nuclear power should not be considered green.

The next step the majority of the funds in my sample (eight out of nine) perform, is positive screening. This means that the funds aim to identify investees which align with the values and objectives of the fund. One example for such a positive screening process is that the fund managers assess the positive impact or the exposure to environmental risks of a business on the transition to a sustainable economy based on sustainability indicators and combine them into a score (e.g., DNCA, 2022, Fund Dact Sheet, 2022; Nordea, 2022a). With these results the funds usually apply a best-in-universe approach and go forward with the investees which were not excluded in the negative screening and had the best score results in the positive screening. Theses scores are not necessarily calculated by the fund provider, but some fund managers also rely on external ratings for this step. Based on these insights I conclude that negative and positive screening are the standard steps which green funds conduct in their selection processes.

I observe more variety in the data collection and analysis processes of green funds within the selection process than for the screening steps. For five of the funds the investment managers are in direct contact with the investees and use this opportunity to gain further insights into the companies' environmental performance and seek confirmation of the performance and actions described in the environmental disclosures for the companies. The fund managers who are in active dialogue with the investees describe this as an integral part of the selection process as it allows them to gain insights into the companies' operational and business practices which go beyond what can be achieved based solely on publicly available data (e.g., Nordea, 2022b; Robeco, 2022c). Four of the fund selection processes are based on extensive in-house research based on which they set up own models and ratings for the selection (e.g., Baillie Gifford, 2022; Mirova, 2022b; Nordea, 2022b; Robeco, 2022b). An intersection of three of these funds is also in active dialogue with the companies and can therefore analyse the additional insights they gain from the contact with the investees and form models and ratings which target their investment objective.

Both, the active dialogue with the investees and the in-house research gives the fund managers the possibility to align their screening and selection better with their objective and value than a selection which is based on only publicly available data. Based on these results I try to provide a categorization of green funds according to their selection processes (see Table 7) as the classification according to SFDR, the investment approach, or the green fund category do not seem to have any patterns in relationship with the selection processes.

The conducting of negative and positive screening seems to be the minimum standard for green funds selection processes based on my results, therefore I class funds which con-

Table 5: Objectives and Investment Approaches of Green Funds

Source: Derived from various sources provided by the fund providers, see Appendix III.
Notes: This table provides an overview of the investment objective of the green funds, the investment approach they apply (see also Table 1), the fund category (see also Table 3) and whether they are an Article 8 or 9 SFDR fund (see also Tables 2 and 4).

		Investment Approach	Green Fund Category	SFDR Article
1	SRI Euro Quality	SRI	Climate Conscious	8
2	ESG Multi-Asset Fund	ESG	Climate Conscious	8
3	Global Impact	Impact Investing	Climate Conscious	8
4	Global SDG Engagement Equities	SRI	Climate Conscious	9
5	Sustainable Energy Fund	ESG	Clean Energy/Tech	9
6	Sustainable Global Thematic Portfolio	ESG	Climate Conscious	9
7	Global Climate and Environment Fund	SRI	Climate Solutions	9
8	Worldwide Positive Change Fund	Impact Investing	Climate Conscious	9
9	Global Sustainable Equity Fund	SRI	Climate Solutions	9

Table 6: Green Fund Selection Process Components

Source: Derived from various sources provided by the fund providers, see Appendix III.
Notes: This table provides an overview of the steps in the fund selection processes and which funds conduct which steps. For a definition of 'Negative Screening' and 'Positive Screening' see Chapter 3.1. 'Active Dialogue' means that the investment manager is in active dialogue with the managers of the investee to gain a better understanding of their business practices and environmental performance. 'In-House Research' describes that the fund providers have a research team dedicated to achieving insights into the environmental performance of the investees and they construct their own models and ratings. (Yes) means that the fund managers do conduct the step, but not with regards to environmental criteria.

		Negative Screening	Positive Screening	Active Dialogue	In-House Research
1	SRI Euro Quality	Yes	Yes	Yes	
2	ESG Multi-Asset Fund	Yes			
3	Global Impact	Yes	Yes	Yes	
4	Global SDG Engagement Equities	Yes	Yes	Yes	Yes
5	Sustainable Energy Fund	Yes	Yes		
6	Sustainable Global Thematic Portfolio	(Yes)	Yes		
7	Global Climate and Environment Fund	Yes	Yes	Yes	Yes
8	Worldwide Positive Change Fund	Yes	Yes	Yes	Yes
9	Global Sustainable Equity Fund	Yes	Yes		Yes

Table 7: A Categorization of Fund Selection Processes

Source: Based on results which are derived from various sources published by the fund providers, see Appendix III.
Notes: This table provides a categorization of the fund selection processes. For an overview of the different selection process components, I identify in my descriptive analysis, see Table 6. These are the basis for 'Requirements'. 'Fund in Category' indicates which funds are sorted into the respective category. For the fund number see Table 6.

Type	Requirements	Funds in Category
1 Basic	Performs one or two of the selection process components.	2, 5, 6
2 Medium	Performs three of the selection process components.	1, 3, 9
3 Advanced	Performs four of the selection process components.	4, 7, 8

duct these two steps as type 1 or 'basic'. Based on my observations the average fund selection process either is in direct dialogue with the investees or conducts in-house research in addition to the negative and positive screening. Thus, I class fund selection processes with three of the components as type 2 or 'medium', which therefore encompasses the average of my sample. Lastly, above average are fund selection processes which incorporate all four of the components. Therefore, I group these as type 3 or 'advanced'. With my sample this results in three green funds per type.

Based on the differences in the selection processes I expect that the holdings of the funds differ depending on the process components. It becomes apparent that some companies were selected by several green funds as illustrated in

Table 8. The table gives an overview of the top ten most selected companies by the green funds and how many times they are selected in the financial year 2021.

Now it would not be surprising that green funds with similar selection processes and similar investment approaches invest in the same companies. One example of the frequently selected companies is ASML Holding NV, a supplier in the semiconductor industry. Based on this example I analyse which funds selected ASML Holding NV as an investee and to which type of investment processes they belong and illustrate the results in Table 9.

This example illustrates that the same company, in this case ASML Holding NV, is selected by funds with different selection processes ranging from the lowest type of selection processes with the only fund of my sample, the ESG Multi-Asset Fund, that just conducts negative screening, to a fund with the most advanced and complex selection process being the Global Climate and Environment Fund. For the other companies in the top ten most selected companies by the green fund in my sample I find a similar picture, a full overview of the top ten companies and the green funds they are selected by see Appendix IV.

Therefore, the differences in the fund selection processes do not allow for conclusions about the selected companies. In my analysis I identify one pattern between the categorization of the selection processes and the result of the selection, which is that funds of the basic type of selection process on average have the most holdings and the funds of the advanced type of selection process have in average the lowest number of holdings in my samples as illustrated in Table 10.

Based on this descriptive analysis of the fund selection processes it becomes evident, that the processes differ from each other and cannot be summarised for green funds in general. Also, based on criteria such as the fund's investment approach, green fund category, or SFDR classification, I cannot make a generalized statement on what their selection process looks like. Therefore, I continue to use the green fund types I deduct based on my observations as presented in Table 7 in my empirical analysis to see whether the association between green funds and environmental disclosure quality, that I expect to find, differs depending on the design of the fund selection process. As some funds are for example in active dialogue with the investees and try to support improvements of environmental performance and disclosures, I expect that the environmental disclosure quality of the companies which are in fund with a more advanced selection process to be different from companies in green funds with more basic selection processes. To understand the relationship of green fund selection processes and environmental disclosure quality I describe the role of environmental disclosure quality in the selection processes of green funds in the next chapter.

5.1.2. Role of Environmental Disclosure Quality in the Selection Processes

Despite the differences in the selection processes of funds, especially regarding the level of detail with which the fund providers conduct their own research on the potential investee, all of the funds mention the disclosures of a company as a relevant source for their selection processes. For an overview on the usage of environmental disclosures of the funds in their selection processes see Table 11.

The documents on three of the funds describe that the environmental disclosures of the firm are the fundament for their assessment of the environmental performance of a company without mentioning further details regarding this process (BlackRock, 2022; Fund Dact Sheet, 2022). To note, two of these three funds are classified as selection process type 1, being the basic selection process type (see Table 7).

The documents provided by the fund managers and providers of six funds discuss more in depth how they use the environmental disclosures within their selection process for own models and ratings (see Table 11, column 'Extended'). Robeco (2022a) for example states for the Global SDG Engagement Equities fund that the information disclosed by a company are an integral part in their assessment of a company as they use them as one of the main data sources in their research center. This research center develops and applies a sustainability framework which defines a baseline which companies have to overcome in order to become part of RobecoSAM's investment universe. According to their framework, the company's disclosures play an integral role in the assessment on whether a firm reaches that baseline and additionally, whether the company should be included in one of their thematic funds. Mirova (2022a) describes that for the Global Sustainable Equity Fund they assess, amongst others, environmental criteria and form a score for the company which later is used to decide whether the fund invests in the company. They state that they gather the main information for the environmental score based on the environmental disclosures of the company. Another example is DNCA (2022). Regarding the SRI Euro Quality fund, they do not only mention environmental disclosures as a fundamental source for their financial and environmental analysis, but also discuss potential difficulties of using environmental disclosures. They state that environmental disclosures are a potentially difficult to use source due to a lack of uniform criteria, definitions, and standards for measurement. Furthermore, they point out that data access and reliability are potential hurdles when utilizing environmental disclosures. But despite these problems, DNCA still concludes that environmental disclosures play a key role in their analysis of companies as they are the only direct and publicly available source of environmental information on a company.

Moreover, four funds explicitly describe that they use external ratings within their selection processes (see column 'Beyond' of Table 11). On the first glance this does not seem to imply a relationship between the selection processes of green funds and environmental disclosures, but the providers of external ratings also mention environmental disclosures as a main source of information in their assessment processes. One by the fund providers frequently mentioned example of such a rating is the MSCI ESG Rating. MSCI (2022) describes in their brochure on the ESG rating that company disclosure

Table 8: Top Ten Selected Companies by the Green Funds

Source: Derived from various sources provided by the fund providers, see Appendix III.
Notes: This table provides an overview of the companies which are the most frequently selected by the green funds in my sample. The funds are presented in Table 6.

Rank	Company	Number of Selections
1	ASML Holding NV	5
2	Orsted AS	4
3	Linde PLC	4
4	Christian Hansen Holding AS	4
5	Infineon AG	4
6	Allianz SE	3
7	Koninklijke DSM NV	3
8	Schneider Electric SE	3
9	Dassault Systems SE	3
10	L'Oreal SA	3

Table 9: Green Funds selecting ASML Holding NV

Source: Derived from various sources provided by the fund providers, see Appendix III.
Notes: This table illustrates which funds selected ASML Holding NV for their portfolio as well as their selection process types (for the types see Table 7). For a similar analysis for all the top ten holdings as presented in Table 8 see Appendix IV.

	Fund	Selection Process Type
1	SRI Euro Quality	2 Medium
2	ESG Multi-Asset Fund	1 Basic
7	Global Climate and Environment Fund	3 Advanced
8	Worldwide Positive Change Fund	3 Advanced
9	Global Sustainable Equity Fund	2 Medium

Table 10: Average Number of Holdings per Green Fund

Source: Derived from various sources provided by the fund providers, see Appendix III.
Notes: This table depicts the average number of holdings which are in my scope per green fund selection process type.

Total Average Number of Holdings: 22	
Type	Average Number of Holdings
1 Basic	36
2 Medium	22
3 Advanced	8

documents are one of their main sources of data for their evaluation of a company. This also encompasses environmental disclosures for the evaluation of the environmental performance of a company. A similar procedure is also described by S&P (2022) within their ESG score methodology, which is another example of a frequently used external source.

The fund providers describe not only the extent of the information presented as important but highlight the importance of their credibility. Therefore, they indirectly describe not only environmental disclosures but also environmental disclosure quality as relevant for their assessment of companies in their selection process, because the extent and credibility are, besides the adoption of environmental disclosures, the components of environmental disclosure quality as discussed in Chapter 2.2.2 and as illustrated by the contributions in prior literature which are presented in Chapter 3.2.

Overall, these examples highlight that despite the differences in the level of detail in which the fund providers discuss their use of environmental disclosures, they all mention them as one of the key sources or even the key source in their environmental analysis of a firm and that the quality of the environmental disclosures is interlinked with their selection of a company. The use of additional external sources such as ratings does not contradict this interlinkage as these external ratings also use environmental disclosures as a main data source in their assessment of the environmental perfor-

Table 11: Environmental Disclosures in the Selection Processes

Source: Derived from various sources provided by the fund providers, see Appendix III.
Notes: This table provides an overview on how the different funds use environmental disclosures in their selection processes. 'Standard' means that the fund managers or providers mention the relevance of environmental disclosures for the selection process, 'Extended' means that they illustrate that they use data out of environmental disclosures in advanced research, ratings and models based on which they make their investment decisions. 'Beyond' means that the fund managers or providers state that they use external ratings within their selection process which again rely on environmental disclosures.

		Standard	Extended	Beyond
1	SRI Euro Quality	Yes	Yes	
2	ESG Multi-Asset Fund	Yes		Yes
3	Global Impact	Yes		
4	Global SDG Engagement Equities	Yes	Yes	Yes
5	Sustainable Energy Fund	Yes		
6	Sustainable Global Thematic Portfolio	Yes	Yes	
7	Global Climate and Environment Fund	Yes	Yes	
8	Worldwide Positive Change Fund	Yes	Yes	Yes
9	Global Sustainable Equity Fund	Yes	Yes	Yes

mance of a company.

5.2. Empirical Analysis of Green Funds and Environmental Disclosure Quality

My descriptive analysis of the fund selection processes reinforces my expectation that the selection by a green fund and the environmental disclosure quality are positively associated and that the environmental disclosure quality increases after the selection by a green fund. Additionally, based on the results from the descriptive analysis, I also expect that the association differs depending on the selection process of the fund. Based on these expectations I present the results of my empirical analysis in the following. The sample on which my analysis is based is illustrated in Table 12.

Based on the descriptive statistics it already becomes visible that companies which are selected for a green fund have an on average higher total environmental disclosure score. Especially the extent and credibility of environmental disclosures are on average higher when compared to the companies which are not selected by a green fund. As the adoption score consists of the general disclosure of environmental information and the adoption of voluntary disclosure standards (see Appendix II), the differences in the adoption score probably stem from differences in the adoption of voluntary disclosure standards, as the general adoption of environmental disclosures became mandatory for all publicly listed firms in the EU with the implementation of the NFRD which was transposed into national law between 2015 and 2018, as discussed in Chapter 2.2.

The correlations of the numeric variables which are presented in Table 13 exhibit increased correlations for the measures of environmental disclosure quality especially among the total environmental disclosure score and the extent as well as credibility of environmental disclosures score. This is reasonable as they are based on the same underlying framework. I investigate them separately as dependent variables in the following regression models.

Table 14 depicts the results of a regression analysis that examines the relationship between the total environmental disclosure score and the selection by a green fund.

The results illustrate that companies that are selected by a green fund have higher environmental disclosure scores than those who are not. The results also provide evidence that environmental disclosures scores of companies that are selected by green funds are higher after the selection. This is as the coefficients of the model including firm fixed effects that are only driven by companies which have variation in the variable green_fund are significantly positive. When controlling for year fixed effects, the results again indicate that the environmental disclosure scores of companies that are selected by green funds are higher than of those who are not. When using both, firm and year fixed effects however, the coefficient becomes negative and insignificant which suggests that the associations observed in the other models are driven by a general time trend and not by the selection by a green fund itself.

An additional investigation of how the association differs for the environmental disclosure quality components, being the adoption, extent, and credibility of the disclosures, does not provide comprehensive additional insights. My findings show a similar pattern when compared to the total environmental disclosure score. The coefficients are significant and positive for the three disclosure components besides when using firm and year fixed effects. The only exemption is the adoption of environmental disclosures of which the coefficient is also insignificant when using only firm fixed effects. This suggests that the adoption of disclosures is not different before and after the selection by a green fund. The implemented regression models that include the components as dependent variables and the selection by a green fund as independent variable are reported in Appendix V.

Moreover, I want to assess whether the association differs depending on the type of green fund by which the companies are selected. For this, I use the green fund types as indepen-

Table 12: Descriptive Statistics (Full Sample)

Source: The presented statistics are based on my sample data. The data is partly obtained from EIKON, partly manually collected. For a description of the data collection see Chapter 4.1.
Notes: This table reports descriptive statistics on the investigated balanced panel sample from 2017 to 2021. For each variable, the number of observations, the mean, the standard deviation, the minimum, the 25th, 50th and 75th percentile as well as the maximum are presented. The treatment group consists of companies which are selected by a green fund, the control group consists of companies which are never selected by a green fund during the sample period. The groups consist of 136 companies each. The variable eds represents the environmental disclosure score and eds_adoption, eds_extent and eds_credibility reflect the three components of the score, being the adoption, extent, and credibility of environmental disclosures (see Chapter 4.1.2 and Appendix II). SIZE, ROA, and LEV are numeric control variables. SIZE is the total assets of the company. ROA is the return on assets and LEV the leverage ratio of the company. SIZE is transformed with the natural logarithm and was originally recorded in thousand EUR. Significance tests show that the control group is significantly different from the treatment group for eds and the score components and not for the control variables SIZE, LEV and ROA. I implement four binary variables which are relevant for the treatment group. green_fund indicates if a company was selected by a green fund in the specific year. Within the treatment group, 97 of 136 companies have variation in the variable green_fund. green_fund type1, green_fund type2 and green_fund type3 indicate whether the company is selected by a green fund type with a selection process of category 1, 2 or 3 in the respective year. For the definition of the green fund selection process categories see Table 7. For the control group the four binary variables take 0 in all observations.

	Mean	S.D.	Min.	25 %	Median	75 %	Max.
Treatment Group n = 680							
eds	40.75	12.58	0.00	34.00	43.50	49.00	68.00
eds_adoption	1.75	0.50	0.00	2.00	2.00	2.00	2.00
eds_extent	27.85	8.82	0.00	23.00	30.00	34.00	47.00
eds_credibility	11.15	4.20	0.00	9.00	12.00	14.00	19.00
SIZE	9.88	1.96	0.00	8.93	9.82	10.99	14.27
ROA	0.06	0.10	-0.50	0.02	0.05	0.08	1.12
LEV	0.21	0.17	0.00	0.10	0.19	0.29	1.77
green_fund	0.63	0.48	0.00	0.00	1.00	1.00	1.00
green_fund type1	0.46	0.50	0.00	0.00	0.00	1.00	1.00
green_fund type2	0.23	0.42	0.00	0.00	0.00	0.00	1.00
green_fund type3	0.11	0.31	0.00	0.00	0.00	0.00	1.00
Control Group n = 680							
eds	31.88	16.98	0.00	22.00	36.00	45.00	64.00
eds_adoption	1.50	0.77	0.00	1.00	2.00	2.00	2.00
eds_extent	21.91	11.64	0.00	15.00	24.00	31.00	43.00
eds_credibility	8.48	5.22	0.00	4.00	9.00	12.00	19.00
SIZE	9.71	2.45	0.00	8.40	10.04	11.14	15.23
ROA	0.05	0.06	-0.23	0.01	0.04	0.08	0.33
LEV	0.21	0.16	0.00	0.08	0.18	0.31	1.00

Table 13: Correlations

Notes: This table reports Pearson correlations above and Spearman correlations below the diagonal for the numeric variables in my sample (n = 1360). For sample and variable definitions see Table 12. Asterisks indicate the significance as follows. * p<0.05/ ** p<0.01/ *** p<0.001

	A	B	C	D	E	F	G
A: eds		0.756***	0.984***	0.918***	0.463***	-0.098***	-0.046
B: adoption	0.599***		0.738***	0.651***	0.316***	-0.095***	-0.035
C: extent	0.971***	0.576***		0.834***	0.441***	-0.083**	-0.055*
D: credibility	0.900***	0.537***	0.779***		0.461***	-0.116***	-0.021
E: SIZE	0.459***	0.255***	0.420***	0.472***		-0.199***	-0.170***
F: ROA	-0.162***	-0.128***	-0.140***	-0.176***	-0.380***		-0.077**
G: LEV	0.006	0.011	-0.004	0.023	-0.095***	-0.070*	

dent variables and the environmental disclosure score and its components as dependent variables. Table 15 reports the results for the total environmental disclosure score.

The results exhibit that the environmental disclosure score has a significantly positive association with all three green fund types when not using fixed effects. The same is

Table 14: OLS Regressions (Total EDS)

Notes: This table reports the results of five OLS regressions with the environmental disclosure score as the dependent variable and the selection by a green fund as independent variable as well as additional control variables. The sample and all variables are defined in Table 12. Standard errors are presented in parentheses below the coefficients. The asterisks indicate the two-sided significance levels and should be interpreted as * p<0.05/ ** p<0.01/ *** p<0.001.

	Dependent Variable:				
	eds (1)	eds (2)	eds (3)	eds (4)	eds (5)
green_fund	9.482***	8.493***	3.565***	7.286***	-0.761
	(0.871)	(0.775)	(0.556)	(0.453)	(0.563)
SIZE		3.141***	3.043***	3.062***	1.224***
		(0.169)	(0.542)	(0.101)	(0.335)
ROA		-2.736	-5.667	-0.908	-0.171
		(4.412)	(5.190)	(0.832)	(4.437)
LEV		2.316	15.018**	1.051	2.108
		(2.241)	(4.883)	(2.195)	(3.184)
Constant	33.311***	2.532			
	(0.491)	(1.904)			
Estimator	ols	ols	ols	ols	ols
Observations	1360	1360	1360	1360	1360
R2	0.080	0.279	0.882	0.302	0.922
R2 Adjusted	0.080	0.277	0.849	0.297	0.900
Fixed Effects	None	None	Firm	Year	Firm, Year

the case when considering firm or year fixed effects. This again indicates that the companies selected by any green fund type have higher environmental disclosure scores than those who are not, and they increase after the selection by any of the three green fund types. The coefficients for green funds of type 1 are higher than of type 2 and 3, which suggests that the environmental disclosure quality of companies selected by green funds of type 1 is higher when compared to types 2 and 3. This could mean that the environmental information disclosed by the investees is less important for funds with more advanced selection processes that use for example active dialogue or in-house research as they do not solely rely on the publicly available information but use additional data sources.

Lastly, I investigate the differences in the associations between the green fund types and environmental disclosure score components. The results of the regression models implemented for this investigation are presented in Tables 16 to 18.

The reported regressions in Table 16 use the environmental disclosure adoption as the dependent variable.

The results in Table 16 indicate that companies selected by green funds of type 1 and 3 have higher adoption scores than those who are not. Companies which have variation in the green fund type variables do not have higher adoption scores before and after the selection by any fund types as the coefficients under consideration of firm fixed effects are insignificant. The differences in the adoption could stem from the adoption of voluntary disclosure standards as the general adoption of environmental disclosures is mandatory for all companies in my sample since the implementation of the NFRD. Table 17 reports the results for regression models with the environmental disclosure extent as dependent variable.

The results in Table 17 indicate that companies which are selected by a green fund of type 1 have a higher extent score than those who are not, this also holds when considering firm or year fixed effects. Therefore, companies which are selected by a green fund of type 1 also have a higher extent after the selection. But again, it seems that this is driven by a general trend as the coefficient in the model with both, firm and year fixed effects, is insignificant. The extent of environmental disclosures of companies selected by a green fund of type 2 is not significantly different in any model. For companies that are selected by a green fund of type 3 the extent score is higher than for those who are not, but it is not higher after the selection than before as indicated by the insignificant coefficient in the model using firm fixed effects. The coefficient for green fund type 1 is higher than for green fund type 3, additional tests however suggest that this difference is not significant. Therefore, the results indicate that the extent of environmental disclosures of companies selected by green funds of type 1 and 3 is higher than for companies which are not selected by these fund types, but only the disclosure extent of companies selected by a type 1 fund increases after the selection.

Table 18 reports the results of regression models using the credibility score as the dependent variable. The results show that the credibility score has a significantly positive associa-

Table 15: OLS Regressions (Total EDS, Fund Types)

Notes: This table reports the results of five OLS regressions with the environmental disclosure score as the dependent variable and the green fund types as independent variables as well as additional control variables. The sample and all variables are defined in Table 12. The green fund types are defined in Table 7. Standard errors are presented in parentheses below the coefficients. The asterisks indicate the two-sided significance levels and should be interpreted as * p<0.05/ ** p<0.01/ *** p<0.001.

	Dependent Variable:				
	eds	eds	eds	eds	eds
	(1)	(2)	(3)	(4)	(5)
green_fund type1	9.424***	8.270***	3.221***	7.600***	-0.407
	(0.980)	(0.873)	(0.614)	(0.228)	(0.564)
green_fund type2	3.158*	2.456*	2.687**	1.097*	-0.957
	(1.298)	(1.152)	(0.921)	(0.328)	(0.930)
green_fund type3	7.165***	7.261***	2.232**	6.575***	-1.570*
	(1.799)	(1.599)	(0.832)	(0.613)	(0.769)
SIZE		3.119***	3.036***	3.035***	1.239***
		(0.169)	(0.545)	(0.094)	(0.339)
ROA		-3.644	-5.727	-1.710*	-0.310
		(4.405)	(5.286)	(0.504)	(4.469)
LEV		2.944	15.132**	1.560	2.151
		(2.240)	(5.112)	(2.025)	(3.175)
Constant	33.423***	2.802			
	(0.479)	(1.900)			
Estimator	ols	ols	ols	ols	ols
Observations	1360	1360	1360	1360	1360
R2	0.088	0.284	0.882	0.310	0.922
R2 Adjusted	0.086	0.281	0.849	0.304	0.900
Fixed Effects	None	None	Firm	Year	Firm, Year

tion with the selection by all three fund types with green fund type 1 having the highest coefficients, even when controlling for firm or year fixed effects. Also, the credibility increases after the selection by all fund types. However, when controlling for both, firm and year fixed effects, the coefficients become negative and insignificant which indicates that despite the credibility scores being higher for companies that are selected by any green fund type, this does not seem to be driven by the selection itself but by a general trend. Additional tests show that the difference in the coefficients is significant when comparing green funds of type 1 and 2 as well as 2 and 3 which suggests that the credibility of environmental disclosures is the highest for green funds of type 1 and the lowest for green funds of type 3. This could indicate that the credibility of disclosures is less relevant for green funds which use active dialogue and in-house research in their selection processes as they have additional sources which they can use to assess the credibility of the information provided by the companies and are therefore not as depending on indicators for credibility which are published by the companies themselves.

Overall, the results of the regression models confirm my expectations that the environmental disclosure quality for green fund holdings is higher. Also, the environmental disclosure quality is increasing after the selection by a green fund. The credibility scores are significantly lower for companies which are selected by green funds with more advanced selection processes. However, the associations do not seem to be driven by the selection itself but by a general trend. With this result in mind, I investigate the development over time in the following.

5.3. Green Funds and Environmental Disclosures over Time

In the following I want to assess how the environmental disclosure scores change over time and whether I find a time trend which is indicated by the empirical results. For a full overview of the environmental disclosure scores and growth rates for the investigated fund types and score components over time see Appendix VI.

Figure 1 shows the development of the total environmental disclosure scores for green fund holdings and non-green fund holdings in comparison. This graph illustrates that the environmental disclosure scores of both groups increase over time but that those of the green fund holdings are always on a higher level over the sample period.

These results illustrate the regression results reported in Table 14 being that the environmental disclosure score is higher for green fund holdings and increasing, but this is

Table 16: OLS Regressions (EDS Adoption, Fund Types)

Notes: This table presents the results of four OLS regressions with the environmental disclosure component adoption as the dependent variable and the green fund types as independent variables as well as additional control variables. The sample and all variables are defined in Table 12. The green fund types are defined in Table 7. Standard errors are presented in parentheses below the coefficients. The asterisks indicate the two-sided significance levels and should be interpreted as * p<0.05/ ** p<0.01/ *** p<0.001.

	Dependent Variable:			
	eds_adoption (1)	eds_adoption (2)	eds_adoption (3)	eds_adoption (4)
green_fund type1	0.158***	0.016	0.145***	-0.063
	(0.041)	(0.031)	(0.014)	(0.035)
green_fund type2	0.085	0.012	0.057*	-0.066
	(0.054)	(0.071)	(0.015)	(0.074)
green_fund type3	0.270***	0.005	0.256**	-0.075*
	(0.075)	(0.031)	(0.050)	(0.038)
SIZE	0.090***	0.125***	0.088***	0.085**
	(0.008)	(0.028)	(0.007)	(0.029)
ROA	-0.324	-0.280	-0.279	-0.149
	(0.206)	(0.266)	(0.143)	(0.262)
LEV	0.067	0.400	0.035	0.103
	(0.105)	(0.210)	(0.070)	(0.179)
Constant	0.686***			
	(0.089)			
Estimator	ols	ols	ols	ols
Observations	1360	1360	1360	1360
R2	0.124	0.793	0.131	0.804
R2 Adjusted	0.120	0.736	0.124	0.749
Fixed Effects	None	Firm	Year	Firm, Year

due to a general trend of increasing environmental disclosure quality for all companies. The scores of the green fund holdings are just on a higher level. This trend is also illustrated by the average and annual compound EDS change of green fund and non-green fund holdings over time which are presented in Table 19. The total score as well as its components increase for both groups and the growth rates are higher for non-green fund holdings.

Figure 2 depicts the development of the total environmental disclosure score before and after the selection by a green fund.

Lastly, I want to investigate the difference in environmental disclosure quality depending on the type of green funds over time. Figure 3 illustrates that the total environmental disclosure scores are increasing for all three fund types over time with the score of holdings of type 1 green funds being the highest. But the scores of types 2 and 3 holdings are catching up over time.

The findings suggests that green funds of type 1 tend to select companies with higher environmental disclosure scores as they are reliant on the published information, but the scores of companies selected by green fund types 2 and 3 increase more over the observed time period. For the total environmental disclosure score, extent and credibility, the average as well as compound annual percentage change is the lowest for fund type 1 and the highest for type 3 as illustrated in Table 20.

Overall, the development of the environmental disclosure scores over time supports the findings that indeed the environmental disclosure quality is higher for companies which are selected by green funds and is increasing after the selection by a green fund. But this seems to be an overall trend of increasing environmental disclosure score as the results also show increasing scores for those companies which are never selected by a green fund. The environmental disclosure scores for companies in green funds are just overall higher. With regards to the differences in selection processes of green fund the results provide evidence that the more sophisticated the fund selection processes, the lower the environmental disclosure scores are in the beginning and the more the environmental disclosure scores of respective fund holdings increase over time.

5.4. Discussion

My findings consistently support my initial expectations that companies in green funds have a higher environmental disclosure quality than companies which are not selected by a green fund as well as that the environmental disclosure

Table 17: OLS Regressions (EDS Extent, Fund Types)

Notes: This table presents the results of four OLS regressions with the environmental disclosure component extent as the dependent variable and the green fund types as independent variables as well as additional control variables. The sample and all variables are defined in Table 12. The green fund types are defined in Table 7. Standard errors are presented in parentheses below the coefficients. The asterisks indicate the two-sided significance levels and should be interpreted as * p<0.05/ ** p<0.01/ *** p<0.001.

	Dependent Variable:			
	eds_extent (1)	eds_extent (2)	eds_extent (3)	eds_extent (4)
green_fund type1	5.578***	1.895***	5.233***	-0.168
	(0.612)	(0.439)	(0.309)	(0.433)
green_fund type2	0.920	0.963	0.245	-1.042
	(0.807)	(0.676)	(0.269)	(0.652)
green_fund type3	5.547***	0.948	5.204***	-1.183*
	(1.120)	(0.509)	(0.280)	(0.542)
SIZE	2.051***	1.870***	2.007***	0.883***
	(0.118)	(0.306)	(0.072)	(0.232)
ROA	-1.256	-3.200	-0.403	-0.637
	(3.086)	(3.800)	(0.404)	(3.405)
LEV	1.353	8.358**	0.677	1.378
	(1.569)	(3.107)	(1.489)	(2.215)
Constant	2.915*			
	(1.331)			
Estimator	ols	ols	ols	ols
Observations	1360	1360	1360	1360
R2	0.260	0.884	0.274	0.910
R2 Adjusted	0.257	0.853	0.268	0.885
Fixed Effects	None	Firm	Year	Firm, Year

quality of companies in green funds increases further after the selection by a green fund. At the same time this does not seem to be a hint for a potential causal influence of the selection by a green fund on the environmental disclosure quality of their investees as my results show a similar trend of an increasing environmental disclosure quality for companies which are not selected by green funds. The environmental disclosure quality though is on a higher level for companies selected by green funds. This could indicate that green funds tend to select companies with a higher environmental disclosure quality which would be in line with the findings in my descriptive analysis which clearly highlight the importance of the information presented in the environmental disclosures of a company for the selection processes of green funds.

For the components of environmental disclosure quality, I find a significant positive association of the selection by a green fund and the adoption, extent as well as the credibility of environmental disclosures. Also, besides for the adoption of environmental disclosures, the quality increases after the selection by a green fund.

When looking at the green fund types my findings show that the selection by a fund of all types is significantly associated with a higher environmental disclosure score. Regarding the credibility my findings illustrate that the scores are lower for funds with more advanced selection processes. This indicates that for funds with advanced selection processes that use active dialogue and in-house research, the credibility of the environmental disclosures is less important as they gain information from additional sources. These results could also indicate that they do not automatically tend to select companies with higher environmental disclosure quality and credibility as they are not reliant on the credibility indicators published by the companies.

With my findings I also want to contribute to the identification of a potential causal relationship between the green fund selection and the environmental disclosure quality of companies. Based on my results I conclude that I would not expect to find a causal influence of the selection by a green fund on the environmental disclosure quality in general. Rather, I assume that the observed increase in environmental disclosure scores is driven by a general trend of increasing environmental disclosure quality. Based on the results of my investigation I suppose the idea that not the selection by a green fund itself but the interaction between green fund managers and the investees, for example due to active dialogue as well as the usage of information gathered directly from the investee in in-house research could have a positive effect on the environmental disclosure quality. This

Table 18: OLS Regressions (EDS Credibility, Fund Types)

Notes: This table presents the results of four OLS regressions with the environmental disclosure component credibility as the dependent variable and the green fund types as independent variables as well as additional control variables. The sample and all variables are defined in Table 12. The green fund types are defined in Table 7. Standard errors are presented in parentheses below the coefficients. The asterisks indicate the two-sided significance levels and should be interpreted as * p<0.05/ ** p<0.01/ *** p<0.001.

	Dependent Variable:			
	eds_credbility (1)	eds_credbility (2)	eds_credbility (3)	eds_credbility (4)
green_fund type1	2.533***	1.310***	2.222***	-0.177
	(0.276)	(0.257)	(0.135)	(0.209)
green_fund type2	1.451***	1.712***	0.795*	0.151
	(0.364)	(0.356)	(0.225)	(0.419)
green_fund type3	1.445**	1.278**	1.115*	-0.312
	(0.505)	(0.465)	(0.296)	(0.364)
SIZE	0.979***	1.041***	0.940***	0.271
	(0.053)	(0.255)	(0.073)	(0.142)
ROA	-2.064	-2.247	-1.028	0.476
	(1.391)	(1.612)	(0.405)	(1.227)
LEV	1.524*	6.374**	0.848	0.669
	(0.707)	(1.979)	(0.546)	(1.030)
Constant	-0.800			
	(0.600)			
Estimator	ols	ols	ols	ols
Observations	1360	1360	1360	1360
R2	0.284	0.833	0.344	0.906
R2 Adjusted	0.281	0.787	0.338	0.880
Fixed Effects	None	Firm	Year	Firm, Year

Table 19: EDS Change - Green Fund vs. Non-Green Fund

Source: Based on my sample data. The sample and all variables are defined in Table 12.
Notes: This table presents the environmental disclosure scores percentage change for the treatment group and control group over time, on average and compound annual. For variable and sample definitions see Table 12. For a full overview of the environmental disclosure scores and growth rates for the investigated fund types and score components over time see Appendix VI.

Year	eds	eds_adoption	eds_extent	eds_credibility
Green Fund (in %)				
Average	6.48	1.92	5.29	10.47
Compound Annual	5.13	1.53	4.19	8.15
Non-Green Fund (in %)				
Average	9.46	6.51	8.16	13.65
Compound Annual	7.48	5.11	6.46	10.64

is supported by the environmental disclosure score growth rates which are considerably higher for green funds which conduct active dialogue, in-house research, or both, than for those who do not. Despite this, my results highlight the overall importance of environmental disclosures for the selection processes of green funds. Furthermore, when thinking about potential causal relationships, I believe that my results also provide insights which hint at a potential causal influence of the environmental disclosure quality on the selection by a green fund, especially for funds of type 1 which do not rely on additional data sources in their selection processes. These trains of thought illustrate interesting questions for future research.

My study is not without uncertainty as there are several limitations to my findings. Firstly, my assessment of environmental disclosure quality is implemented in form of a score

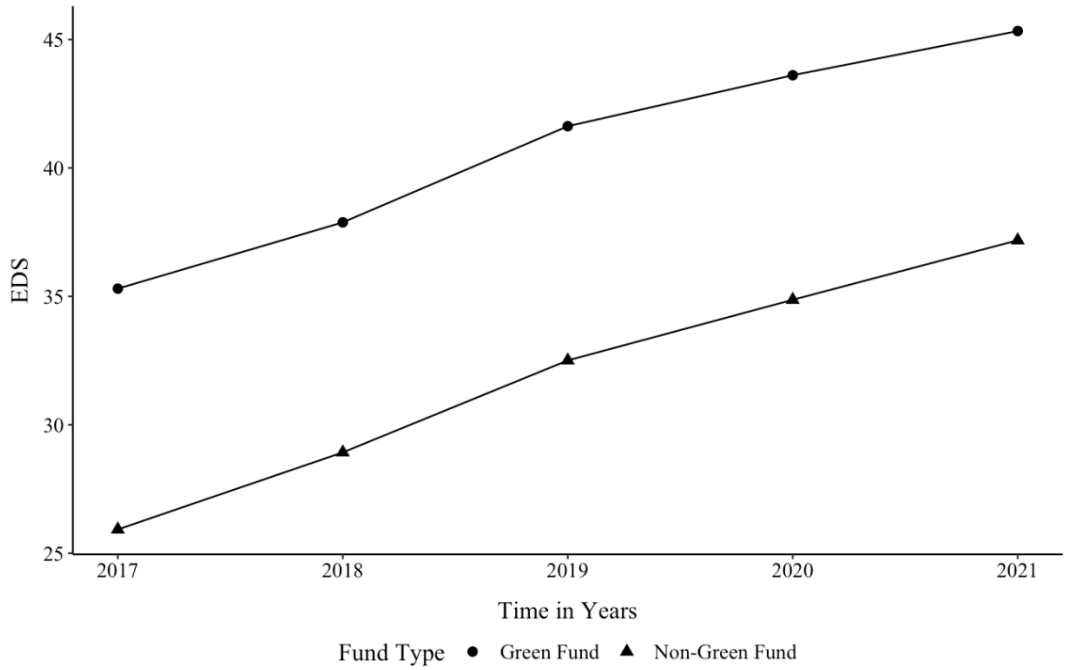


Figure 1: Total EDS over Time – Green vs. Non-Green Funds

Source: Based on the full sample data, the sample and all variables are defined in Table 12.
Notes: This graph depicts the total environmental disclosure score data and compares the scores of companies which are selected by a green fund to those that are not selected by a green fund during the sample period.

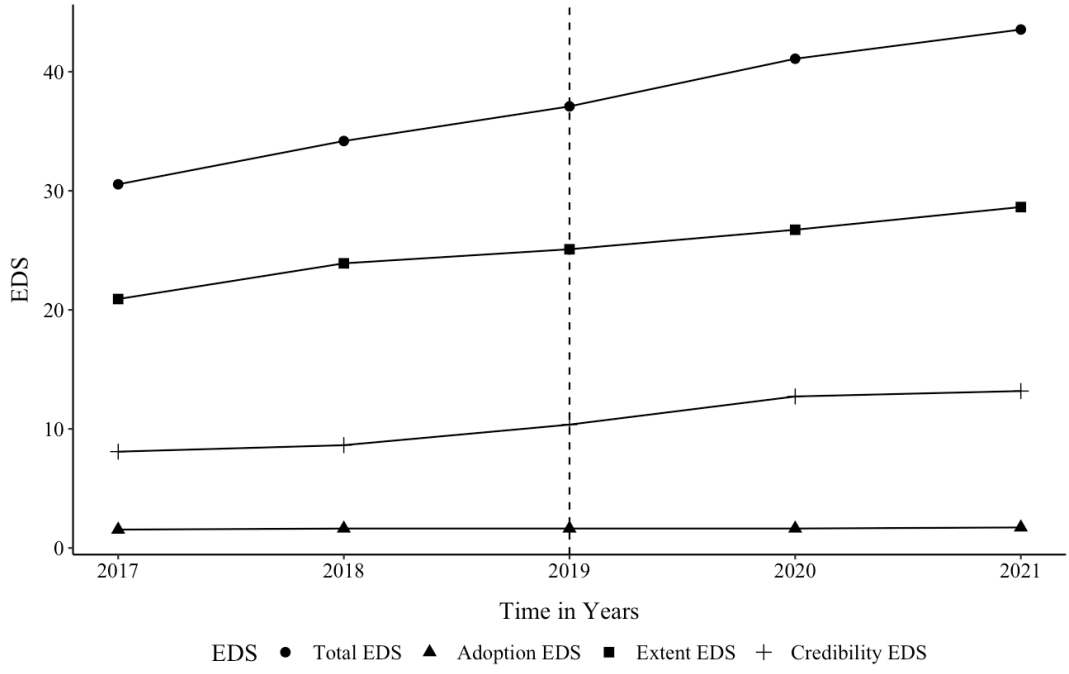


Figure 2: EDS – Before and After Green Fund Selection

Source: Based on treatment group data, the sample and all variables are defined in Table 12.
Notes: This graph depicts the environmental disclosure score data for companies which were initially selected by a green fund in 2019. This is illustrated by the dashed line. It shows that the scores do indeed increase further after the selection as also seen in the empirical results in Chapter 5.2, but I do not observe a change different from the growth before. This suggests that this further increase is not driven by the selection for a green fund itself but by the observed trend of increasing environmental disclosure quality.

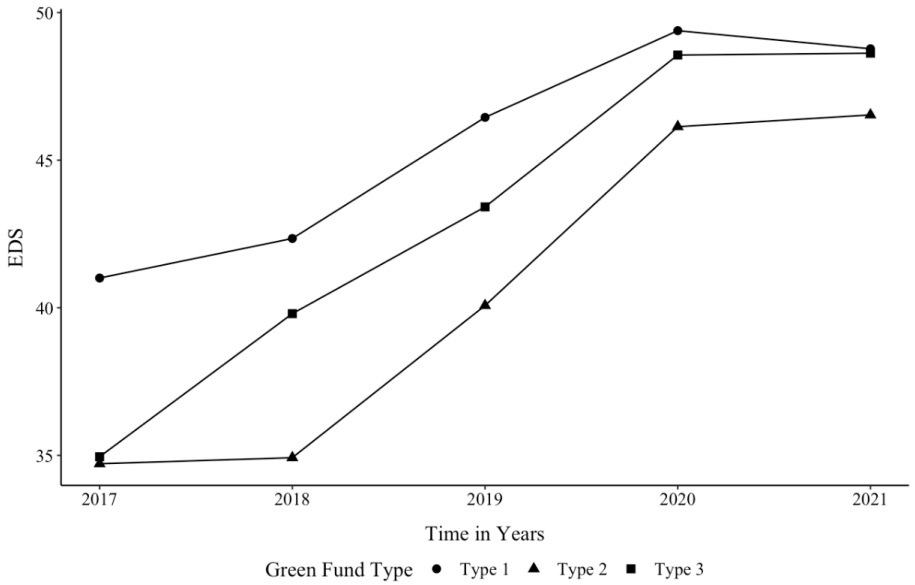


Figure 3: Total EDS over Time – Green Fund Types

Source: Based on treatment group data. The sample and all variables are defined in Table 12.
Notes: This graph depicts the total environmental disclosure score data and compares the scores of companies which are selected by the different types of green funds.

Table 20: EDS Change - Green Fund Types

Source: Based on treatment group data. The sample and all variables are defined in Table 12.
Notes: This table presents the environmental disclosure scores percentage change for the treatment group split up by fund type over time, on average and compound annual. For the fund type definitions see Table 7. For a full overview of the environmental disclosure scores and growth rates for the investigated fund types and score components over time see Appendix VI.

Year	eds	eds_adoption	eds_extent	eds_credibility
Type 1 (in %)				
Average	4.51	1.06	1.63	14.47
Compound Annual	3.53	0.84	1.29	10.70
Type 2 (in %)				
Average	7.83	0.13	5.18	17.98
Compound Annual	6.04	0.07	4.08	12.81
Type 3 (in %)				
Average	8.73	-0.18	6.41	19.56
Compound Annual	6.83	-0.15	4.96	14.31

which can never fully cover all aspects of environmental disclosure quality, as also outlined in the discussion on scores in prior literature illustrated in Chapter 3.2. Secondly, the outcome of the score is affected by the data quality in the database and despite the control of several samples I cannot ensure that the data is fully correct or unbiased. Moreover, I base my analysis on selected green funds and conventional funds of which I identify the holdings. Even though I cover different as well as large fund providers, this does not mean that my analysis covers the complete variety of fund selection processes. Lastly, my findings focus on the situation in the EU. Therefore, the findings might not apply in other regions.

6. Conclusion

This analysis explores the relationship between the selection of companies by a green fund and their environmental disclosure quality. The findings show that the environmental disclosure quality of companies which are selected by green funds is higher than of those who are not selected. Additionally, I find that the environmental disclosure quality of

companies increases after they are selected by a green fund. But my further investigations of this development hint that this is not triggered by the selection itself, but the environmental disclosure quality just increases further. This further increase is not specific to the holdings of green funds but I observe a general trend of increasing environmental disclosure quality for all companies in my sample. Furthermore, I find that the relationship of green fund selection and environmental disclosure quality differs depending on the design of the selection processes of green funds. The environmental disclosure quality of holdings of green funds with basic selection processes is initially higher, but the environmental disclosure quality of holdings of funds with more advanced selection processes catches up quickly. Moreover, my findings suggest that funds with more advanced selection processes who use additional data sources within the selection processes seem to be less dependent on the environmental disclosures, especially regarding the credibility.

My results provide several queries which offer potential for interesting future research. As discussed, I find a general trend of increasing environmental disclosure quality over all companies. This could encourage an investigation of whether this trend is caused by for example the increasing regulatory pressure in the EU where mandatory environmental disclosures are already implemented and standards for the disclosures are planned to be implemented in the next years. Another potential explanation would be that I observe this increasing trend because all the companies in my sample are selected by an investment fund, green or conventional. The first idea could be tested based on my sample data with an inferential analysis of the changes in environmental disclose quality after the implementation of the environmental disclosure regulation in the EU whilst the other would require a larger sample which uses companies, which are not a holding of any fund as the control group. Furthermore, an investigating of a potential causal effect of the active dialogue or in-house research of green funds on the environmental disclosure scores can be conducted based on my sample under isolation of the funds which conduct these steps and could potentially explain the differences in the association between green funds and environmental disclosures depending on the green fund selection processes. Also, the data sample allows for an investigation of a potential causal influence of the environmental disclosure quality on the selection by a green fund and whether such a potential effect could be weaker or stronger depending on how much the specific fund relies on environmental disclosures as a data source in their selection process. Furthermore, literature and media currently discuss extensively whether green funds are actually green. My results indicate that green funds do invest in companies with a higher environmental disclosure quality which, if seen as an indicator for environmental performance as discussed in prior literature would mean that green funds do indeed invest in companies which are greener than the investees of conventional funds. Therefore, revisiting the relationship of environmental disclosures and environmental performance in this context could provide important insights for this discussion.

Overall, my analysis provides a detailed description of green fund selection processes and how these selection processes interact with the environmental disclosures of companies. The empirical results highlight the relationship between the selection by a green fund and the environmental disclosure quality of companies. By this I add to literature as I initially investigate this relationship not only theoretically but with a structured descriptive as well as empirical analysis. Moreover, I provide further insights into a topic which is currently of high timeliness but still lacks systematic investigations. Therefore, a structured investigation of green funds in the universe of sustainable finance contributes to the overall understanding of this vast field. Finally, the highlighted trend of an overall increasing environmental disclosure quality can be seen as a silver lining in the debate regarding whether environmental disclosures might just be a platform for greenwashing, especially due to the relevance of credibility of disclosures in the selection processes of green funds as well as the high increases of the credibility scores for both, holdings of green and non-green funds.

References

Arribas, I., Espinós-Vañó, M. D., García García, F., & Oliver-Muncharaz, J. (2019). Defining socially responsible companies according to retail investors' preferences. *Enterpreneurship and Sustainability Issues*, *7*(2), 1641–1653.

Arribas, I., Espinós-Vañó, M. D., García García, F., & Tamosiuniene, R. (2019). Negative screening and sustainable portfolio diversification. *Enterpreneurship and Sustainability Issues*, *6*(4), 1566–1586.

Baillie Gifford. (2022). *Baillie gifford worldwide funds prospectus 2022.* Retrieved from https://www.bailliegifford.com/en/germany/professional-investor/literature-library/funds/dublin-funds/fund-literature/baillie-gifford-worldwide-funds-prospectus/ (Last Access: July 8, 2022)

Bancel, F., & Mittoo, U. R. (2009). Why do european firms go public? *European financial management*, *15*(4), 844–884.

Bassen, A., Gödker, K., Lüdeke-Freund, F., & Oll, J. (2019). Climate information in retail investors' decision-making: Evidence from a choice experiment. *Organization & Environment*, *32*(1), 62–82.

Birkey, R. N., Michelon, G., Patten, D. M., & Sankara, J. (2016). Does assurance on csr reporting enhance environmental reputation? an examination in the us context. In *Accounting forum* (Vol. 40, pp. 143–152).

BlackRock. (2022). *EU taxonomy implementation 2022.* Retrieved from https://www.blackrock.com/corporate/literature/prospectus/eu-taxonomy.pdf (Last Access: July 8, 2022)

Braam, G. J., de Weerd, L. U., Hauck, M., & Huijbregts, M. A. (2016). Determinants of corporate environmental reporting: The importance of environmental performance and assurance. *Journal of cleaner production*, *129*, 724–734.

Busch, T., Bauer, R., & Orlitzky, M. (2016). Sustainable development and financial markets: Old paths and new avenues. *Business & Society*, *55*(3), 303–329.

Campbell, D., Craven, B., & Shrives, P. (2003). Voluntary social reporting in three ftse sectors: a comment on perception and legitimacy. *Accounting, Auditing & Accountability Journal*.

Capelle-Blancard, G., & Monjon, S. (2012). Trends in the literature on socially responsible investment: Looking for the keys under the lampost. *Business ethics: a European review*, *21*(3), 239–250.

Chatterji, A. K., Levine, D. I., & Toffel, M. W. (2009). How well do social ratings actually measure corporate social responsibility? *Journal of Economics & Management Strategy*, *18*(1), 125–169.

Chen, L., Srinidhi, B., Tsang, A., & Yu, W. (2016). Audited financial reporting and voluntary disclosure of corporate social responsibility (csr) reports. *Journal of Management Accounting Research*, *28*(2), 53–76.

Clarkin, J. E., & L. Cangioni, C. (2016). Impact investing: A primer and review of the literature. *Entrepreneurship Research Journal*, *6*(2), 135–173.

Clarkson, P. M., Li, Y., Richardson, G. D., & Vasvari, F. P. (2008). Revisiting the relation between environmental performance and environmental disclosure: An empirical analysis. *Accounting, organizations and society*, *33*(4-5), 303–327.

Clarkson, P. M., Overell, M. B., & Chapple, L. (2011). Environmental reporting and its relation to corporate environmental performance. *Abacus*, *47*(1), 27–60.

Del Giudice, A., & Rigamonti, S. (2020). Does audit improve the quality of esg scores? evidence from corporate misconduct. *Sustainability*, *12*(14), 5670.

Diez-Cañamero, B., Bishara, T., Otegi-Olaso, J. R., Minguez, R., & Fernández, J. M. (2020). Measurement of corporate social responsibility: A review of corporate sustainability indexes, rankings and ratings. *Sustainability*, *12*(5), 2153.

DNCA. (2022). *DNCA SRI euro quality monthly reporting as of may 2022.* Retrieved from https://www.dnca-investments.com/funds/dnca-sri-euro-quality/parts/nc-fr0013514882/MR-nc-fr0013514882/download_doc_share?locale=en&country_code=INT (Last Access: July 8, 2022)

EFRAG. (2022). *Draft european sustainability reporting standard 1 general provisions.* Retrieved from https://www.efrag.org/Assets/Download?assetUrl=%2fsites%2fwebpublishing%2fSiteAssets%2fWorking%2520Paper%2520ESRS%25201%2520General%2520Provisions.pdf (Last Access: July 1, 2022)

Elkington, J. (1998). Accounting for the triple bottom line. *Measuring business excellence*.

ESMA. (2022). *Mifid ii, policy activities.* Retrieved from https://www.esma.europa.eu/policy-rules/mifid-ii-and-mifir (Last Access: August 3, 2022)

European Commission. (2011). *Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions: A renewed eu strategy 2011-14 for corporate social responsibility.* Official Journal of the European Union, Brussels, Belgium.

European Commission. (2017). *Guidelines on non-financial reporting: methodology for reporting non-financial information.* Official Journal of the European Union, Brussels, Belgium.

European Commission. (2019). *Guidelines on non-financial reporting: Supplement on reporting climate-related information.* Official Journal of the European Union, Brussels, Belgium.

European Commission. (2020). *Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions: Stepping up europe's 2030 climate ambition - investing in a climate-neutral future for the benefit of our people.* Official Journal of the European Union, Brussels, Belgium.

European Parliament. (2014). *Directive 2014/95/eu of the european parliament and of the council of 22 october 2014 amending directive 2013/34/eu as regards disclosure of non-financial and diversity information by certain large undertakings and groups.* Official Journal of the European Union, Brussels, Belgium.

European Parliament. (2019). *Regulation (eu) 2019/2088 of the european parliament and of the council of 27 november 2019 on sustainability-related disclosures in the financial services sector.* Official Journal of the European Union, Brussels, Belgium.

European Parliament. (2020). *Regulation (eu) 2020/852 of the european parliament and of the council of 18 june 2020 on the establishment of a framework to facilitate sustainable investment, and amending regulation (eu) 2019/2088.* Official Journal of the European Union, Brussels, Belgium.

European Parliament. (2021). *Proposal for a directive of the european parliament and of the council amending directive 2013/34/eu, directive 2004/109/ec, directive 2006/43/ec and regulation (eu) no 537/2014, as regards corporate sustainability reporting.* Official Journal of the European Union, Brussels, Belgium.

Fernandez-Feijoo, B., Romero, S., & Ruiz, S. (2014). Effect of stakeholders' pressure on transparency of sustainability reports within the gri framework. *Journal of business ethics*, *122*, 53–63.

Fernando, S., & Lawrence, S. (2014). A theoretical framework for csr practices: Integrating legitimacy theory, stakeholder theory and institutional theory. *Journal of Theoretical Accounting Research*, *10*(1), 149–178.

Foster, N. J. (2021). Our commitment to ESG principles. *The Risk Management Association Journal*, *7*(103), 3.

Fund Dact Sheet. (2022). *Sustainability Policy 2022.* Retrieved from https://www.ubs.com/2/e/files/RET/FS_RET_LU1679116845_LI_EN.pdf (Last Access: July 8, 2022)

Gangi, F., D'Angelo, E., et al. (2016). The virtuous circle of corporate social performance and corporate social disclosure. *Modern Economy*, *7*(12), 1396.

GRI. (2022). *Sustainability reporting guidelines.* GRI Universal Standard, Amsterdam, The Netherlands.

Hahn, R., & Kühnen, M. (2013). Determinants of sustainability reporting: A review of results, trends, theory, and opportunities in an expanding field of research. *Journal of cleaner production*, *59*, 5–21.

Hellsten, S., & Mallin, C. (2006). Are 'ethical'or 'socially responsible'investments socially responsible? *Journal of Business Ethics*, *66*, 393–406.

Hoffmann, J., Scherhorn, G., & Busch, T. (2004). Darmstadt definition of sustainable investments. *Retrieved from Wuppertal: https://www.econs tor. eu/bitst ream/10419/59280/1/48495*, *7511*.

Ibikunle, G., & Steffen, T. (2017). European green mutual fund performance: A comparative analysis with their conventional and black peers. *Journal of Business Ethics*, *145*, 337–355.

ICMA. (2021). *Green bond principles - voluntary process guidelines for issuing*

green bonds. Retrieved from https://www.icmagroup.org/assets/documents/Sustainablefinance/2021-updates/Green-Bond-Principles-June-2021-140621.pdf (Last Access: June 17, 2022)

Inderst, G., Kaminker, C., & Stewart, F. (2014). *Defining and measuring green investments: Implications for institutional investors asset allocations* (Tech. Rep.). OECD Working Papers on Finance, Insurance and Private Pensions.

ISSB. (2022). *Draft ifrs s1 general requirements for disclosure of sustainability-related financial information.* Retrieved from https://www.ifrs.org/content/dam/ifrs/project/general-sustainability-related-disclosures/exposure-draft-ifrs-s1-general-requirements-for-disclosure-of-sustainability-related-financial-information.pdf (Last Access: July , 2022)

Ito, Y., Managi, S., & Matsuda, A. (2013). Performances of socially responsible investment and environmentally friendly funds. *Journal of the Operational Research Society*, *64*(11), 1583–1594.

Kaufer, K., & Steponaitis, L. (2021). *Just money: Mission-driven banks and the future of finance*. MIT Press.

Kolk, A., & Perego, P. (2010). Determinants of the adoption of sustainability assurance statements: An international investigation. *Business strategy and the environment*, *19*(3), 182–198.

KPMG. (2022). *The kpmg survey of sustainability reporting 2020, kpmg impact, 11th edition.* Retrieved from https://home.kpmg/xx/en/home/insights/2020/11/the-time-has-come-survey-of-sustainability-reporting.html (Last Access: July 02, 2022)

Lagasio, V., & Cucari, N. (2019). Corporate governance and environmental social governance disclosure: A meta-analytical review. *Corporate Social Responsibility and Environmental Management*, *26*(4), 701–711.

Lozano, R., & Huisingh, D. (2011). Inter-linking issues and dimensions in sustainability reporting. *Journal of cleaner production*, *19*(2-3), 99–107.

Maltais, A., & Nykvist, B. (2020). Understanding the role of green bonds in advancing sustainability. *Journal of sustainable finance & investment*, 1–20.

Michelon, G., Pilonato, S., & Ricceri, F. (2015). Csr reporting practices and the quality of disclosure: An empirical analysis. *Critical perspectives on accounting*, *33*, 59–78.

Mirova. (2022a). *Mirova minimum standards for controversial activities 2021.* Retrieved from https://www.mirova.com/sites/default/files/2021-01/Controversial-Activities-Jan-2021-EN.pdf (Last Access: July 8, 2022)

Mirova. (2022b). *Mirova our approach to esg assessment 2021.* Retrieved from https://www.mirova.com/sites/default/files/2021-03/our-approach-to-esg-assessment.pdf (Last Access: July 8, 2022)

Morningstar. (2022a). *Global sustainable fund flows: Q1 2022 in review, morningstar manager research.* Retrieved from https://www.morningstar.com/lp/global-esg-flows (Last Access: June 17, 2022)

Morningstar. (2022b). *SFDR article 8 and article 9 funds: 2021 in review, morningstar manager research.* Retrieved from https://www.morningstar.com/lp/sfdr-article8-article9 (Last Access: July 11, 2022)

MSCI. (2022). *MSCI ESG rating brochure, MSCI ESG research.* Retrieved from https://www.msci.com/documents/1296102/21901542/MSCI+ESG+Ratings+Brochure-cbr-en.pdf (Last Access: July 11, 2022)

Nordea. (2022a). *Fund fact sheet 2022.* Retrieved from https://www.nordea.de/de/professional/documents/monthly-report---fact-sheet/MR_N1_GCLIMEF_BP_EUR_ger_INT.pdf (Last Access: July 8, 2022)

Nordea. (2022b). *Nordeas commitment to the european sri transparency code 2022.* Retrieved from https://www.nordea.de/de/professional/documents/static-links/2022_Transparency-Code-40_ENG_Equity_NAM_ESG_STARS_and_ESG_Thematics.pdf (Last Access: July 8, 2022)

Pérez-López, D., Moreno-Romero, A., & Barkemeyer, R. (2015). Exploring the relationship between sustainability reporting and sustainability management practices. *Business Strategy and the Environment*, *24*(8), 720–734.

Robeco. (2022a). *Fund factsheet 2022.* Retrieved from https://www.robeco.com/doca/CGF_SDGEE_DH-fact-202205-profdede.pdf (Last Access: July 8, 2022)

Robeco. (2022b). *Pre-contractual disclosure of a fund that has a sustainable investment objective 2022.* Retrieved from https://www.robeco.com/docm/sfdr-global-sdg-engagement-equities.pdf (Last Access: July 8, 2022)

Robeco. (2022c). *SDG framework 2022.* Retrieved from https://www.robeco.com/docm/docu-robeco-explanation-sdg-framework.pdf (Last Access: July 8, 2022)

Ryall, C., & Riley, S. (1996). Appraisal of the selection criteria used in green investment funds. *Business Strategy and the Environment*, *5*(4), 231–241.

SASB. (2020). *Conceptual framework exposure draft.* Retrieved from https://www.sasb.org/wp-content/uploads/2021/07/PCP-package_vF.pdf (Last Access: July 1, 2022)

Semenova, N., & Hassel, L. G. (2015). On the validity of environmental performance metrics. *Journal of business ethics*, *132*, 249–258.

S&P. (2022). *S&P global ESG scores methodology, S&P global.* Retrieved from https://www.spglobal.com/esg/documents/sp-global-esg-scores-methodology-2022.pdf (Last Access: July 11, 2022)

Stuart, E., & Bioy, H. (2021). *Investing in times of climate change, morningstar manager research.* Retrieved from https://www.morningstar.com/en-uk/lp/investing-in-timesofclimatechange?utm_medium=referral&utmcampaign=linkshare&utm_source=link (Last Access: June 17, 2022)

TCFD. (2017). *Final report - recommendations of the task force on climate-related financial disclosures.* Retrieved from https://assets.bbhub.io/company/sites/60/2021/10/FINAL-2017-TCFD-Report.pdf (Last Access: July 1, 2022)

TCFD. (2018). *Task force on climate-related financial disclosures: status report.* Retrieved from www.fsb-tcfd.org/wp-content/uploads/2018/08/FINAL-2018-TCFD-Status-Report-092518.pdf (Last Access: July 1, 2022)

Windolph, S. E. (2011). Assessing corporate sustainability through ratings: challenges and their causes. *Journal of Environmental sustainability*, *1*(1), 5.

Junior Management Science 8(3) (2023) 798-826

Junior Management Science

journal homepage: www.jums.academy

Blockchain Technology Adoption among Consumers: An Analysis of Usage Intention and Application Usefulness

Dennis Henning

Technische Universität München

Abstract

Blockchain technology research has mainly been focused on general usage intention, mostly examined the organizational perspective, and lacked a differentiated view at specific blockchain applications from the consumer perspective. To foster adoption of blockchain technology, consumer perception of blockchain technology needs further understanding. Building on recent technology adoption literature and employing a representative survey for Germany, we identified context dependent predictors and moderators of blockchain technology usage intention. Results show that drivers of usage intention depend on consumers' age, gender, experience, and cryptocurrency possession. Findings guide practitioners by shedding light on blockchain adoption and usefulness of specific blockchain applications. Moreover, results indicate that blockchain adoption research should be more granular and differentiate between applications and contexts. Our identified specific blockchain applications provide a basis for future research.

Keywords: Blockchain technology; Technology adoption; UTAUT; Usage intention.

1. Introduction[1]

Numerous practitioners and scholars believe that blockchain technology has the potential to disrupt many industries and to be a main force in modern businesses (Aydiner, 2021; Chong, Lim, Hua, Zheng, & Tan, 2019; Cong & He, 2019; Frizzo-Barker et al., 2020; Weking et al., 2020). In fact, venture funding to blockchain startups surged by 713% from 2020 to 2021 to reach $25.2 billion, while the number of global blockchain unicorns increased by 422% from nine in 2020 to 47 in 2021 (CB Insights, 2021). However, as our studies of the German and British consumer market show, only 3% and 6% of consumers (e.g., end-users) have knowingly used blockchain applications so far, respectively (see also Knauer & Mann, 2020). These findings contrast the current hype around blockchain technology and raise questions about the underlying drivers of blockchain usage intention and specific application usefulness perceptions from a consumer's perspective. This paper attempts to give answers within the fields of technology adoption by providing an empirical analysis.

At the intersection of computer science, cryptography, and economics, blockchain is thought to be a foundational technology of the fourth industrial revolution (Iansiti & Lakhani, 2017; Toufaily, Zalan, & Dhaou, 2021). At its core, blockchain refers to a decentralized ledger technology that enables serial, peer-to-peer transactions without third-party intermediaries (Liang, Kohli, Huang, & Li, 2021; Toufaily et al., 2021). Its key characteristics constitute anonymity, transparency, security, traceability, and efficiency of transactions (Liang et al., 2021). This allows for publicly auditable ledgers that simultaneously preserve the privacy of the individual (Yin, Langenheldt, Harlev, Mukkamala, & Vatrapu, 2019).

Numerous studies indicate that blockchain technology has the potential to create value in several ways (Abdollahi, Sadeghvaziri, & Rejeb, 2022; Nowiński & Kozma, 2017; Weking et al., 2020; Zheng & Boh, 2021). First, blockchain technology creates an ecosystem of actors that removes the need for a third party to establish trust between participants (Ali, Jaradat, Kulakli, & Abuhalimeh, 2021; Rossi, Mueller-Bloch, Thatcher, & Beck, 2019; Weking et al., 2020;

[1]This thesis is based on the forthcoming paper by Mehrwald & Henning (2022): Consumers' Perspective on Blockchain Technology: What drives Usage Intention and determines Application Usefulness?

DOI: https://doi.org/10.5282/jums/v8i3pp798-826

Zhang, Wei, Jiang, Peng, & Zhao, 2021). Instead, trust is established among all parties through immutable and transparent transactions as well as validated records (Weking et al., 2020; Zhang et al., 2021). Therefore, blockchain offers users a decentralized mechanism for authenticating data and transactions, setting it apart from centralized transaction systems (Weking et al., 2020). Second, blockchain creates cost reduction potentials that allow users to benefit from lower transaction costs, e.g. in financial payments (Abdollahi et al., 2022; Nowiński & Kozma, 2017). Lower transaction costs emerge from disintermediation, reduced record-keeping for customers and faster transaction times improving operational efficiency of businesses, as well as enhanced data traceability and verification (Nowiński & Kozma, 2017; Weking et al., 2020; Zheng & Boh, 2021). Third, blockchain could create societal enrichments through democratization, new business practices and extended access domains (e.g., new financial resources, crowdsourcing, new stakeholders) (Abdollahi et al., 2022).

Initially popularized as the technology behind the cryptocurrency Bitcoin (Cong & He, 2019), blockchain has been increasingly utilized as a building block for a wide range of use cases in many different domains (Marikyan, Papagiannidis, Rana, & Ranjan, 2022). Currently, sectors like finance, supply chain management, healthcare, voting, arts and entertainment witness a strong interest in blockchain use cases (Ali et al., 2021). Those use cases mostly build on the following blockchain applications: Self-sovereign identity, tokenization, fractional ownership, micropayments, smart contracts and (pseudo-)anonymous transactions. The literature revealed those applications to be main drivers for new business models (Boston Consulting Group, 2019; Schlecht, Schneider, & Buchwald, 2021; Toufaily et al., 2021; Zheng & Boh, 2021; Ziolkowski, Miscione, & Schwabe, 2020).

As a top technology trend, blockchain needs to be widely adopted and diffused if the innovation is to realize its socioeconomic benefits (Toufaily et al., 2021). Regardless of its benefits, value drivers or applications, widespread adoption is still rare. To foster adoption, consumers' perception of blockchain technology needs further understanding. This is important, because consumers are a decisive factor for the long-term success of blockchain technology applications. According to Toufaily et al. (2021), consumers are expected to reap benefits from more efficient transactions (e. g., inexpensive and fast payments), increased transparency, verifiability and accuracy of information, as well as self-sovereign data ownership and identity control. However, consumers are challenged by the technological complexity of blockchains (Marikyan et al., 2022). They find it difficult to understand its services, benefits and use cases, not to mention the technical nuances of its infrastructural layer (Marikyan et al., 2022).

Yet, many researchers have studied the adoption of blockchain technology from an organizational perspective (Liang et al., 2021; Toufaily et al., 2021) or have analyzed its technical advantages and values (Li, 2020). However, empirical research from the perspective of the consumer is still scarce. Particularly, too little attention has been given to the influences of consumers' blockchain usage intention and consumers' assessment of blockchain application usefulness. This differs from the organizational perspective fundamentally because 1) organizations often focus on incrementally improving or digitizing current partnerships and dataflows therein with blockchain, 2) organizations might risk some of their current advantages concerning the use of customer data and intermediating services once blockchain becomes widely used, and 3) business-to-business use cases often remain for a longer period and have more interactions compared to what is relevant among consumers.

Understanding the consumer perspective is then important for the following reasons. First, a consumer's usage intention is a prerequisite for actual usage (Venkatesh, Morris, Davis, & Davis, 2003). Thus, identifying influencing factors for usage intention is necessary to drive actual usage of blockchain technology; for example, by adequately communicating and addressing those influencing factors. Second, studying business models targeted at consumers need a granular level of understanding of which blockchain technology applications potentially address user needs, e.g., microtransactions or rather tokenized assets. This allows focusing researchers' and practitioners' efforts on more specific, useful aspects of blockchain technology. Third, blockchain technology is a decentralized technology and consumers will most likely continue to be the most essential user segment. In this paper, we attempt to fill this gap of lacking consumer focused blockchain research and aim at offering an answer to the following research question: *What influences blockchain usage intention from a consumer perspective and which blockchain applications are considered the most useful by consumers?*

In particular, this study investigates:

1. potential predictors and moderators of blockchain technology usage intention according to recent academic literature (Blut, Chong, Tsigna, & Venkatesh, 2022);
2. which predictors affect consumers' blockchain usage intention;
3. application usefulness of specific blockchain technology applications, namely self-sovereign identity, tokenization, fractional ownership, micropayments, smart contracts and (pseudo-)anonymous transactions.

We build upon the stream of research on technology adoption, like the state-of-the-art and revised unified theory of acceptance and use of technology (UTAUT; Venkatesh et al., 2003) by Blut et al. (2022), and the stream of research focused on the potential of blockchain technology for organizations, consumers, and business models. To answer our research question, we employ a three-step approach. First, we conduct a systematic literature review on predictors of blockchain technology acceptance following Webster & Watson's (2002) guidelines. We use our findings to extend the UTAUT by predictors relevant to our context (Blut et al., 2022) and derive hypotheses. Second, we conduct

a quantitative survey, which is representative for the German population. Third, we statistically examine consumers' intention to use blockchain technology as well as their associated usefulness for identified blockchain technology applications. Our identified predictors include elements of the Technology Readiness Index (TRI), consisting of optimism, innovativeness, discomfort, and insecurity (Parasuraman, 2000; Parasuraman & Colby, 2015). Also, we include context specific predictors, such as social influence, disposition to privacy, trust, perceived risk, perceived benefit for society, potential of disruption and perceived usefulness. We examine these variables by developing two research models that test the effects of the identified variables on blockchain usage intention and application usefulness. Moreover, we test for moderation effects related to usage intention by age, gender, experience, and possession of cryptocurrency (Blut et al., 2022). Specifically, we conduct a (moderated) multiple regression analysis (Hair, 2014) based on our survey (N = 847) in Germany.

Our results are presented in three dimensions, namely predictors for intention to use, moderation effects influencing intention to use and predictors of usefulness for certain blockchain applications. Our results on usage intention reveal that innovativeness, trust, and perceived usefulness have a positive effect on usage intention. Discomfort and perceived risk are found to have a negative and social influence to have a positive effect on usage intention, in the models including gender, experience, or possession of cryptocurrency. A positive relationship is observed for potential of disruption and usage intention, in the models including age, gender, or possession of cryptocurrency. No effect is confirmed for optimism, insecurity, disposition to privacy and perceived benefit for society. Regarding moderation effects, we observe 1) age to negatively affect the relationship between trust and usage intention, 2) gender (males) to negatively influence the effect of perceived risk on usage intention, 3) experience to positively affect the relationship between trust and usage intention as well as to negatively affect the relationship between perceived usefulness and usage intention, and 4) possession of cryptocurrency to positively influence the relationship between trust and usage intention as well as perceived risk and usage intention. Our results on application usefulness show that trust and potential of disruption have a positive effect for every application of our sample. Optimism and perceived benefit for society are found to positively influence application usefulness as well, except for micropayments. Social influence has a positive effect for tokenization and fractional ownership applications, disposition to privacy a negative effect for self-sovereign identity and smart contract applications. No significant relationship is observed for innovativeness, discomfort, insecurity, and perceived risk. To provide additional descriptive value of our sample, we perform a latent-class analysis (LCA) based on the TRI item scores (Parasuraman & Colby, 2015). Results depict the technology readiness and affinity of the sample population. We identify 15% of German respondents to be associated as Explorers, 36% as Pioneers, 28% as Hesitators and 21% as Avoiders.

This paper makes several contributions to theory as well as to practice. First, this is one of the first papers to identify and investigate the drivers of blockchain usage intention from the perspective of the consumer by combining streams of technology adoption literature. Our results refine current UTAUT-, TRI-, and blockchain specific theory and reveal which predictors are relevant in the context of blockchain adoption. Second, this research shows the relevance of including individual characteristics and context specific moderators, such as possession of cryptocurrency. Third, as called for by Rossi et al. (2019), our findings reveal which specific applications might be most promising from the perspective of the consumer. Fourth, we demonstrate which factors organizations should address to influence adoption. Lastly, we foster contextualization in technology adoption research by providing a status quo on the perception of blockchain technology by consumers in Germany and the United Kingdom (UK) as well as a cluster analysis based on the technology readiness of the German and British population. Our study guides further research to a more differentiable view at blockchain applications and calls on examining those which consumers find useful.

The following sections of this paper are structured as follows: We begin by presenting a field report of the perception of blockchain technology by consumers in Germany and the UK as well as their technology readiness. Next, we provide an overview of blockchain technology and technology adoption research based on our systematic literature review. Thereafter, we derive hypotheses and design the research model. We continue by setting out the methodology of our research, covering analysis, measures, as well as reliability and validity assessments. Subsequently, we present the results of our analysis. We combine our findings with the insights gained from literature by providing theoretical and practical implications in the discussion. We point out limitations and give an outlook on future research. We conclude this paper by giving a summary of our work.

2. Field report

2.1. Consumer perception of blockchain technology in Germany and the United Kingdom

Before investigating the influences of blockchain usage intention and analyzing the usefulness of specific blockchain applications, we examine the status quo on consumer awareness and perception of blockchain technology in Germany and the UK. For that purpose, we conducted two surveys: One for the German and one for the British population. Data on Germany was collected via the fieldwork agency Consumerfieldwork, an online research panel service provider. The survey was live for eleven days in October 2021. Data on the UK was gathered via the online research panel service provider Prolific. This survey was live for eleven days in February 2022. During data cleaning, we excluded those respondents who failed age or attentiveness checks to account

for common method bias (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003). This resulted in a final set of N = 847 observations for Germany and N = 898 observations for the UK. Table 1 provides an overview of the baseline characteristics of the German sample population in this study and Table 2 the characteristics of the British sample population.

A comparison of the sample baseline characteristics indicates that German consumers (G) seem to be less aware of the terms associated with blockchain than British consumers (B) ("Blockchain technology": G = 46.6%, B = 62.8%; "Ethereum": G = 32.6%, B = 52.8%; "NFT" (Non-Fungible Token): G = 17.5%, B = 61.9%). The following observations can be made about the set of respondents who have heard about blockchain technology: I. Most came across this term in the domain of finance and banking (G = 41%, B = 51%); II. Relatively more British consumers have heard the term in the sector of arts and collectibles than German consumers (G = 7%, B = 20%); III. Few associations of the term were made in the domains of transport and logistics (G = 9%, B = 5%), energy and utilities (G = 9%, B = 4%), and healthcare and pharmaceuticals (G = 6%, B = 4%); IV. Few German and British consumers have knowingly used blockchain applications so far (G = 7%, B = 10%).

Cryptocurrencies and NFTs still receive little attention from consumers overall. On the one hand, cryptocurrencies have only been possessed by 12% of German consumers and NFTs by 1%. Interestingly, segmenting by gender reveals that 17% of men and 7% of women in Germany have already owned cryptocurrencies. British consumers, on the other hand, show higher possession rates, specifically 27% have owned cryptocurrencies and 4% NFTs. 38% of men and 17% of women in Britain have at some point in their lives possessed cryptocurrencies. VISA (2021) reports similar cryptocurrency possession levels and a male skewness for Britain.

The following observations can be made about the share of respondents who possess(ed) cryptocurrencies: I. Both samples show that these consumers find it relatively easy to purchase cryptocurrencies (mean for G = 2.8, mean for B = 2.9; scale from 1 (*very easy*) to 7 (*very hard*)); II. Coinbase and Binance are the primary exchanges with which the majority of consumers manage their cryptocurrencies (G = 62%, B = 67%), followed by MetaMask or other digital wallets (G = 20%, B = 20%).

Furthermore, out of all respondents, 27% of German and 47% of British consumers would use cryptocurrency as means of payment at some point in the future. The most frequently mentioned reason for which cryptocurrencies would not be used for payment purposes is a lack of interest in cryptocurrencies among the relevant consumers (G = 72%, B = 57%).

When asked about their knowledge of blockchain technology, German and British consumers show rather low levels (mean for G = 2.6, mean for B = 2.5; scale from 1 (*no knowledge*) to 10 (*expert knowledge*)). Differentiating by gender reveals that men have slightly more knowledge than women (mean for men = 3, mean for women = 2). Out of German consumers, 25% know the difference between Bitcoin and blockchain technology. Specifically, 18% of the German sample population are both male and know the difference and only 7% are both female and know the difference. As for Britain, 37% know the difference. Segmenting the British sample population by gender reveals that 26% are both male and know the difference, whereas only 11% are female and know the difference.

Consumers' exposure to blockchain technology and its usage is still very limited. Yet, they have slightly more contact with blockchain technology in their private lives than in their professional lives (combined mean of contact in personal life = 1.8, combined mean of contact in professional life = 1.3; scale from 1 (*very low*) to 7 (*very high*)). When asked about whether they would use blockchain technology, only 20% of German consumers and 30% of British consumers answered "Yes". See Figure 1 for a comparison of consumers' usage intentions.

Consumers' awareness of blockchain technology applications and their engagement with it is still low (See Figure 2). Additionally, consumers feel rather discouraged by their circle of friends to use blockchain technology (mean for G = 3.3, mean for B = 3.5; scale from 1 (*they would discourage me*) to 10 (*they would encourage me*)).

Consumers in Germany and Britain indicate restrained behavior in situations that reflect the functional traits of blockchain technology. For instance, only 5% of German consumers and 2% of British consumers would put their bank account statement on the street in a hypothetical scenario, where everyone could view the statement, but the consumer's name is removed and just their bank account number, transaction data and account balance remain. Both consumer groups feel rather neutral towards the fact that with blockchain technology, their personal details are public, but encrypted as a string of numbers and letters (e.g., "39XpoaixBAbUZzaq7g7"), which ensures that their identity is not revealed (mean for G = 3.6, mean for B = 3.9; scale from 1 (*not comfortable at all*) to 7 (*very comfortable*)). Yet only 38% of Germans and 47% of Brits would transfer money to a verified seller without a name, but just a string of numbers, for the online purchase of an item of medium value (e.g., Bluetooth speaker). Both consumer groups show slight privacy concerns when using blockchain technology for financial transactions (mean for G = 4.5, mean for B = 4.3; scale from 1 (*no privacy concerns*) to 7 (*strong privacy concerns*)). For purchasing a pizza, only 20% of German consumers would use blockchain technology (B = 32%), 18% would use it to buy a jacket (B = 30%), 15% to buy a car (B = 19%) and 13% to buy a house (B = 15%). This could be a descriptive indication that with increasing monetary value, the intention of consumers to purchase via blockchain technology seems to decrease.

Consumers show rather low levels of trust in blockchain technology and its users. Figure 3 provides an overview of consumers' trust in other blockchain users. Figure 4 depicts consumers' trust in blockchain's integrity, benevolence and ability (Hawlitschek, Teubner, & Weinhardt, 2016). When asked about their general disposition to trust, which is a per-

Table 1: German sample baseline characteristics.

Gender	n	%	Age	n	%	Education	n	%	Employment	n	%	Heard of...[a]	n	%
Female	429	50.6	16-34	224	26.5	Incomplete High School	132	15.6	Full-time employee	315	37.2	Bitcoin	837	98.8
Male	418	49.4	35-49	190	22.4	Apprenticeship	253	29.9	Part-time employee	127	15.0	Cryptocurrency	833	98.3
			50-64	238	28.1	High School	157	18.5	Self-employed	41	4.8	Blockchain technology	395	46.6
			65-79	195	23.0	Bachelor	124	14.6	Unemployed/ retired	264	31.2	Ethereum	276	32.6
						Master	137	16.2	Temporary job	21	2.5	NFT	148	17.5
						Other	44	5.2	Other	79	9.3			
Total	847	100.0	Total	847	100.0	Total	847	100.0	Total	847	100.0			

Note: [a] Reflects the number and percentage of participants responding "yes" to this question.

Table 2: British sample baseline characteristics.

Gender	n	%	Age	n	%	Education	N	%	Employment	n	%	Heard of…[a]	n	%
Female	465	51.8	16-34	258	28.7	Incomplete High School	31	3.4	Full-time employee	357	39.8	Bitcoin	896	99.8
Male	433	48.2	35-49	253	28.2	Apprenticeship	42	4.7	Part-time employee	141	15.7	Cryptocurrency	888	98.9
			50-64	286	31.8	High School	273	30.4	Self-employed	93	10.4	Blockchain technology	564	62.8
			65-79	101	11.3	Bachelor	355	39.5	Unemployed/ retired	215	23.9	Ethereum	474	52.8
						Master	146	16.3	Temporary job	6	0.6	NFT	556	61.9
						Other	51	5.7	Other	86	9.6			
Total	898	100.0	Total	898	100.0	Total	898	100.0	Total	898	100.0			

Note: [a] Reflects the number and percentage of participants responding “yes” to this question.

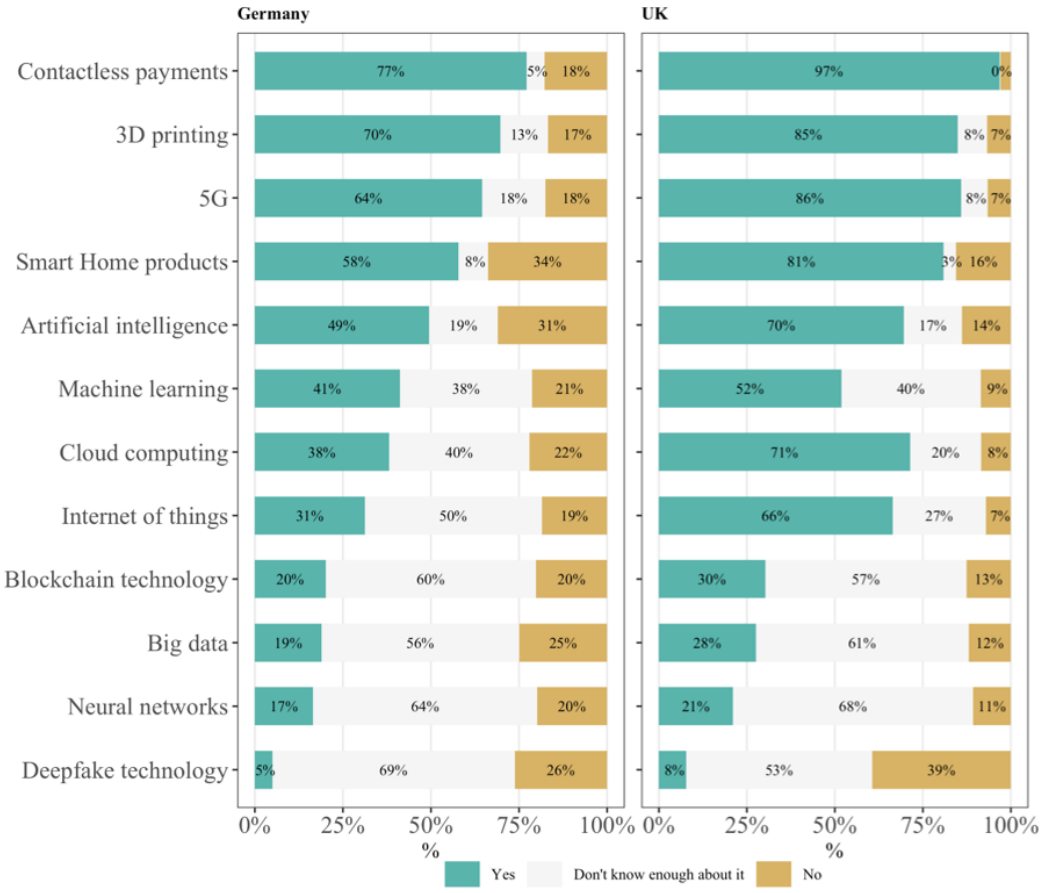


Figure 1: Consumers' usage intention towards different technologies.

Note: $N_{GER} = 847$. $N_{UK} = 898$. The question asked the participant whether they would, purely intuitively and given the chance, use the mentioned technologies.

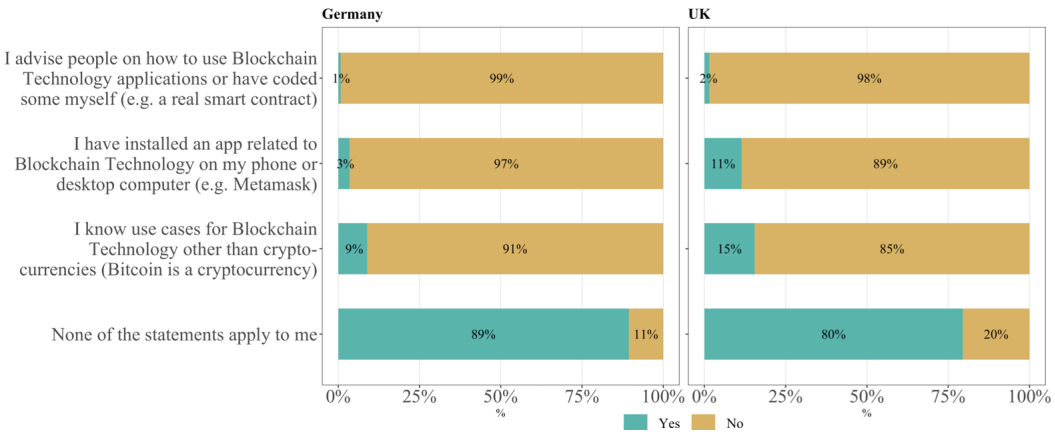


Figure 2: Consumers' awareness of blockchain technology applications.

Note: $N_{GER} = 847$. $N_{UK} = 898$.

son's general inclination to display faith in humanity and to adopt a trusting stance towards others (Gefen, 2000), 66% of German participants and 60% of British respondents answered "You cannot be careful enough".

In sum, it is not only the case that the overall awareness of blockchain technology, its applications, cryptocurrencies and NFTs, is quite low, but also that consumers in Germany and the UK demonstrate a rather cautious attitude towards

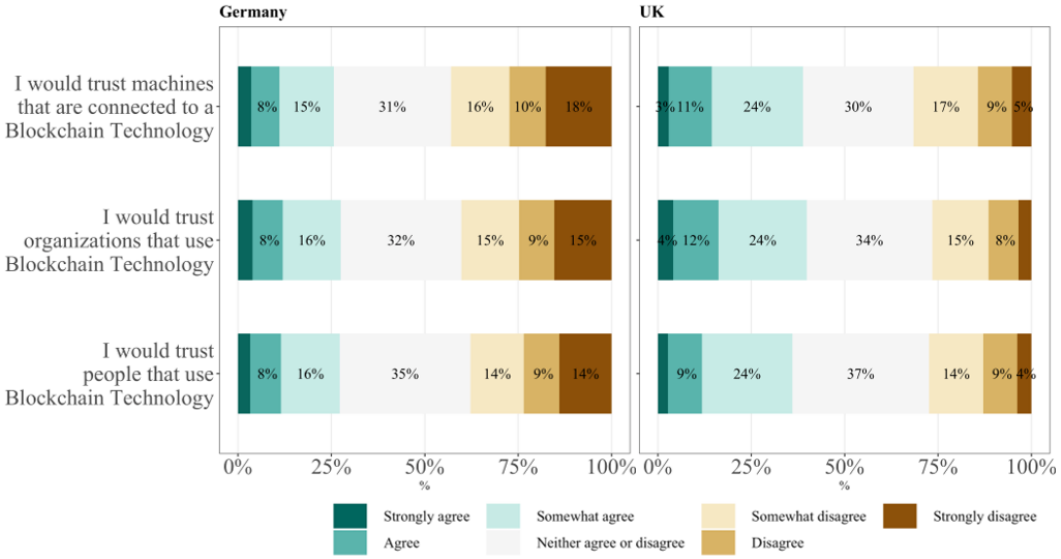


Figure 3: Consumers' trust in other blockchain technology users.

Note: $N_{GER} = 847$. $N_{UK} = 898$.

blockchain technology. The lack of knowledge and trust on part of the consumers could be obstacles, which weigh in on the limited usage intention and adoption of blockchain technology. However, British consumers seem to hold a slightly more approving attitude towards blockchain technology than German consumers. Nevertheless, from the perspective of the consumer, blockchain technology is still perceived to be in its infancy.

2.2. Consumer technology readiness in Germany and the United Kingdom

To better understand people's propensity to embrace and use cutting-edge technologies, we implemented the Technology Readiness Index (TRI) in our surveys (Parasuraman & Colby, 2015). In the TRI, two motivational and two inhibitory forces are considered, which collectively determine a person's predisposition to use new technologies (Parasuraman, 2000). Motivators are the drivers that improve a person's technology readiness, which comprise of optimism – a person's positive view of technology – and innovativeness – a person's willingness to try out new technology (Agarwal & Prasad, 1998; Blut & Wang, 2020; Parasuraman, 2000). Inhibitors are the detractors that lower an individual's technology readiness, which entail discomfort – a person's perceived lack of control over technology and a feeling of being overwhelmed by it – and insecurity – a person's distrust of technology and skepticism about its ability to work properly (Blut & Wang, 2020; Parasuraman, 2000). Extant research shows that higher levels of technology readiness are correlated with higher adoption rates of cutting-edge technology, more intense usage of technology and greater perceived ease in doing so (Parasuraman & Colby, 2015).

Technology readiness is measured using an abbreviated version of TRI 2.0 in our study. The abbreviated index is comprised of ten items[2] covering the four abovementioned constructs, whereby each item is measured on a seven-point Likert scale, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Figure 5 provides an overview of the answers of both sample populations on the TRI items.

Operationalizing technology readiness by applying the TRI allows us to segment our German and British samples into distinct clusters of technology-related beliefs (Parasuraman & Colby, 2015). For that purpose, we conducted a latent class analysis (LCA) (Magidson & Vermunt, 2004) of the TRI item scores. Due to 25 invalid answers for the German sample and 18 for the British sample, the sample size for the cluster analysis had to be reduced to $N_{GER} = 822$ and $N_{UK} = 880$.

The LCA of the German sample population's responses on TRI items results in four clusters of general technology readiness. A comparison of the Bayesian Information Criteria (BIC) of a three-, four-, five- or six-cluster solution demonstrates best fit for the four-cluster solution as indicated by the lowest BIC score ($BIC_3 = 28088.98$, $BIC_4 = 28032.88$, $BIC_5 = 28104.56$, $BIC_6 = 28264.16$) (Magidson & Vermunt, 2004). Moreover, the four-cluster solution demonstrates better distinguishability between the clusters as opposed to the five-cluster solution by Parasuraman and Colby (2015). To maintain comparability of results, the four-cluster solution is applied for the British sample population as well.

We classify 15% (125) of the German sample population

[2]The initial development of the TRI 1.0 is based on 36 items, whereas its updated version, TRI 2.0, is reduced to a 16-item scale (Parasuraman & Colby, 2015). For our purposes, we implemented an abbreviated TRI 2.0 index of ten items, as this version is also capable of predicting TR segment membership with a high degree of accuracy while leaving room for other questions in the survey (see also https://rockresearch.com/abbreviated-version-tri-2-0/).

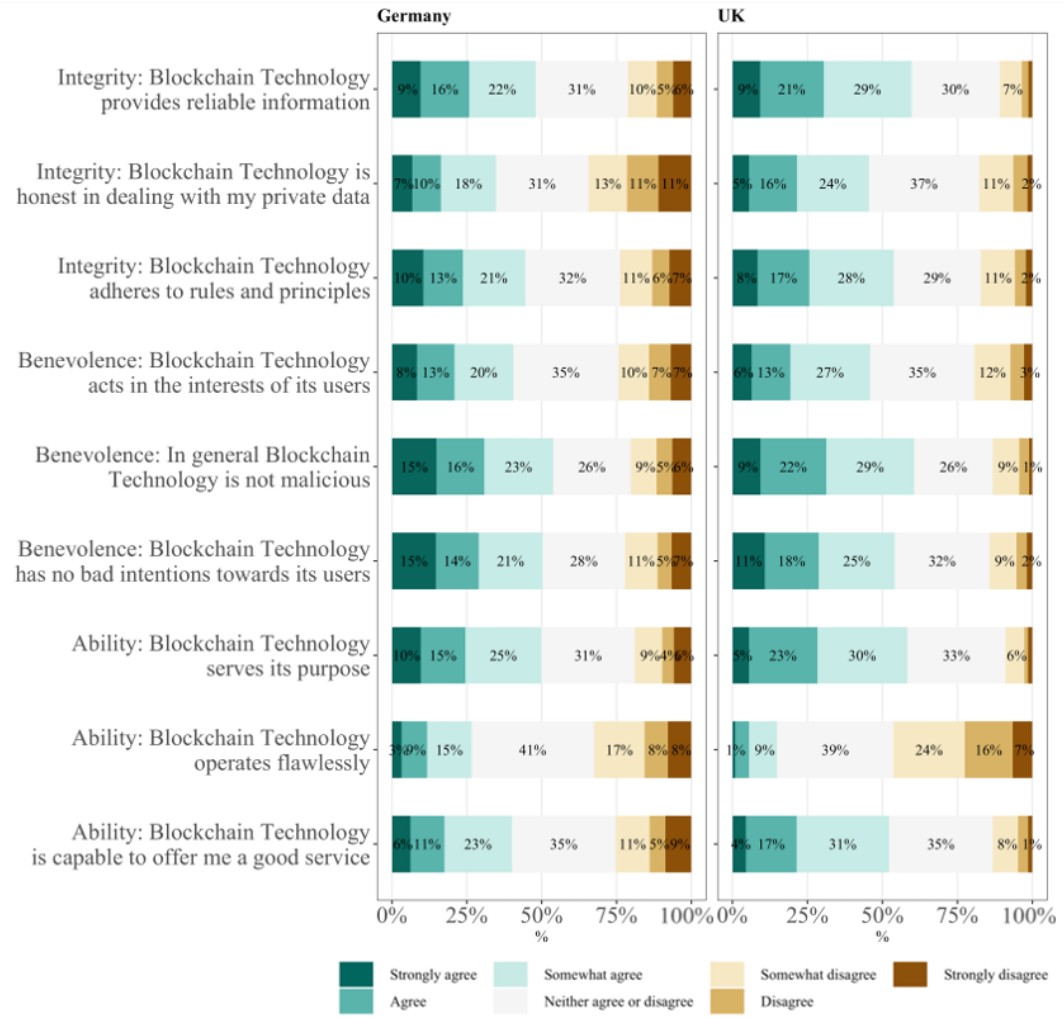


Figure 4: Consumers' trust in blockchain's integrity, benevolence, and ability.

Note: $N_{GER} = 847$. $N_{UK} = 898$.

as Explorers, 36% (297) as Pioneers, 28% (227) as Hesitators and 21% (173) as Avoiders. As for Britain, 13% (114) of respondents are considered Explorers, 26% (226) Pioneers, 40% (352) Hesitators and 21% (188) Avoiders. Following Parasuraman and Colby (2015), Explorers are key consumers or lead users who have a strong motivation to use technology (highest optimism and innovativeness scores) while having a low degree of resistance (lowest discomfort and insecurity scores). Pioneers tend to hold both rather strong positive and negative technology-related beliefs. Hesitators have a high degree of resistance as well as a particularly low degree of innovativeness. Avoiders show the highest degree of resistance and lowest degree of motivation. Referring these clusters to Rogers' (1962) classifications in his theory on diffusion of innovations, Explorers are similar to innovators and early adopters, Pioneers are related to the early majority, Hesitators are similar to the late majority and Avoiders are related to laggards. Table 3 and Table 4 display a summary of the TRI-based LCA results of the German and British sample population, respectively.

British consumers show a stronger technology affinity than German consumers. The mean TRI score of the British sample population is 4.22, whereas the mean TRI score of the German sample population is 3.75. The British sample population reveals stronger motivational forces across clusters while levels of discomfort and insecurity are lower. Thus, although the LCA reveals 40% of the British sample population to be Hesitators, their level of motivation is much higher while inhibitory levels are lower than the corresponding levels in the German Hesitator cluster. Additionally, note that insecurity levels of British consumers appear to be much lower than for German consumers. This might be an indication that German consumers have stronger safety concerns and tend to expect risks rather than benefits in a technology.

The four clusters of technology readiness have distinct demographic and technology-related characteristics (see Table 5 and Table 6). For instance, the cluster with the highest technology readiness, the Explorers, consists of relatively

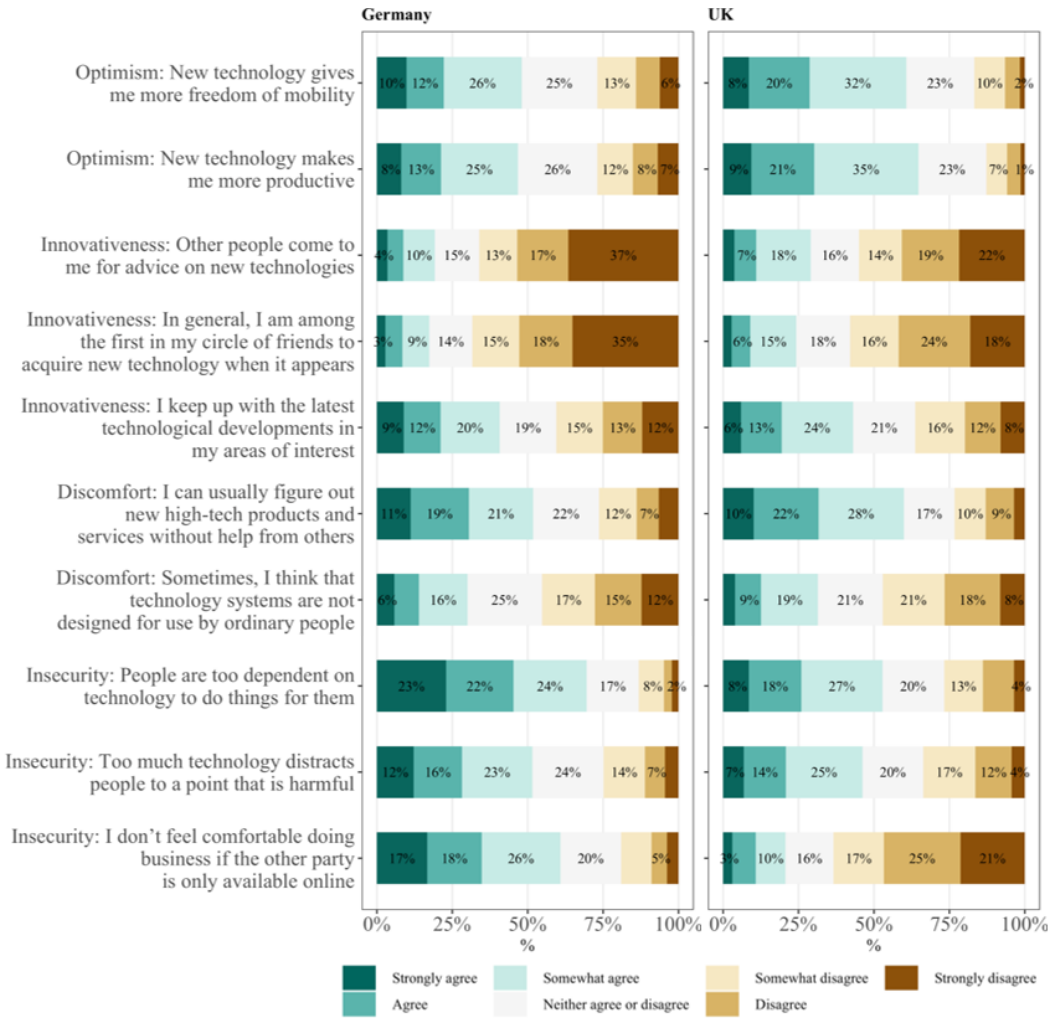


Figure 5: Consumers' technology readiness.

Note: $N_{GER} = 847$. $N_{UK} = 898$.

Table 3: Latent class segmentation using TRI data of German sample population.

Cluster	n	%	Optimism	Innovativeness	Discomfort	Insecurity	Overall TRI
Explorers	125	15	6.04	5.47	2.38	4.06	5.27
Pioneers	297	36	4.82	3.86	3.18	4.74	4.19
Hesitators	227	28	3.78	2.07	3.97	4.51	3.34
Avoiders	173	21	2.81	1.61	4.60	6.03	2.45

Note: $N_{GER} = 822$.

more men, is more highly educated, and possesses comparatively more knowledge about blockchain technology or the internet. Pioneers are even younger but have slightly less technology related knowledge, which applies especially for German consumers. Both Explorers and Pioneers possess more cryptocurrencies and NFTs than the other clusters, which could be a descriptive indication that technology adoption might be higher for Explorers and Pioneers, as suggested by the literature (Parasuraman & Colby, 2015). Avoiders constitute the polar opposite to the Explorers and Hesitators stand in between Pioneers and Avoiders in terms of cluster characteristics.

Table 4: Latent class segmentation using TRI data of British sample population.

Cluster	n	%	Optimism	Innovativeness	Discomfort	Insecurity	Overall TRI
Explorers	114	13	6.06	5.34	2.09	2.62	5.67
Pioneers	226	26	5.20	4.84	3.03	3.85	4.79
Hesitators	352	40	4.62	3.05	3.65	3.90	4.03
Avoiders	188	21	3.80	1.58	4.76	4.69	2.98

Note: $N_{UK} = 880$.

Table 5: Demographic and technology characteristics of German TRI-based clusters.

Cluster	Female (%)	Age 50+ (%)	Min. Bachelor's degree (%)	Knowledge BT[1]	Explain BT[1]	Explain the Internet[1]	Possess. of cryptocurr. (%)	Possess. of NFT (%)	Know diff. between Bitcoin & BT (%)
Explorers	32	51	41	3,95	2,72	5,51	22	5	50
Pioneers	42	43	37	3,07	2,18	5,04	18	1	35
Hesitators	65	52	30	1,83	1,33	4,22	4	-	10
Avoiders	61	62	21	1,62	1,25	4,02	5	-	6

Note: $N_{GER} = 822$. BT = blockchain technology. [1]Question is measured on a scale from 1 (*no knowledge/ do not know how it works*) to 10 (*expert knowledge/ fully capable to explain how it works*).

Table 6: Demographic and technology characteristics of British TRI-based clusters.

Cluster	Female (%)	Age 50+ (%)	Min. Bachelor's degree (%)	Knowledge BT[1]	Explain BT[1]	Explain the Internet[1]	Possess. of cryptocurr. (%)	Possess. of NFT (%)	Know diff. between Bitcoin & BT (%)
Explorers	32	36	63	3.46	2.55	5.11	39	5	57
Pioneers	34	34	59	3.33	2.47	5.17	42	9	54
Hesitators	59	42	61	2.16	1.57	4.30	22	2	32
Avoiders	74	61	53	1.59	1.16	3.75	13	1	12

Note: $N_{UK} = 880$. BT = blockchain technology. [1]Question is measured on a scale from 1 (*no knowledge/ do not know how it works*) to 10 (*expert knowledge/ fully capable to explain how it works*).

3. Literature review

To evaluate the current state of research on technology adoption and on the potentials of blockchain technology, we conducted a systematic literature review according to the guidelines by Webster and Watson (2002). For our search, we used the EBSCOhost Business Source Complete database. To ensure for high-quality scientific knowledge in the field of information system, we searched seven of the eight journals of the Senior Scholars' Basket of Journals[3]. We also included the 50 journals in the Financial Times 50 List in our search to explore the broader implications of technology adoption and blockchain for organizations, consumers, and business models. The initial search had a three-dimensional keyword design: the first field of research covered the keywords "blockchain", "business model" and "distributed ledger technology", the second field of research "industry", "application" and "perception", the third "potential", "innovation", "opportunity", "transformation", "impact", "use" and "usage". We further aimed to focus our review on the latest scientific research by restricting our search to the time frame from 2016 to 2022. We specified the language to be English. This resulted in 153 articles that were eligible for review. After examining the titles and abstracts regarding the fit of the articles for this paper, 30 articles were chosen for a full text analysis. Two articles had to be discarded as their full text was not available (such as "Call for Papers"). After reviewing

[3]EBSCOhost Business Source Complete does not provide access to the Journal of Strategic Information Systems.

the 28 remaining articles, we identified additional 20 articles during forward and backward search. Thus, a total of 48 articles were considered for the literature review. Moreover, to enhance the practical merit of our paper, we complemented our literature review with grey literature. This comprises, for instance, reports by consultancies (Boston Consulting Group, 2019), newspaper articles (Quiroz-Gutierrez, 2022) as well as insights on blockchain technology from market intelligence platforms (Amberdata, 2022). See Appendix 1 for an overview of the literature review methodology.

3.1. Blockchain technology and its applications

Blockchain technology is a decentralized ledger that allows tamper-proof, transparent storage of data and enables peer-to-peer transactions without a central party (Liang et al., 2021; Nakamoto, 2008; Toufaily et al., 2021; Yin et al., 2019). Blocks of transactions are saved and stored in nodes that are encrypted using pseudonyms and are only known to the parties to the transactions (Liang et al., 2021). Therefore, a system of accountability is enabled, while not revealing a user's true identity (Raddatz, Coyne, Menard, & Crossler, 2021; Yin et al., 2019).

The cryptocurrency Bitcoin was the first application for blockchain technology (Nakamoto, 2008) and more than 13,000 cryptocurrencies have been established since then (CoinGecko, 2022). Further developments of blockchain technology expanded the possibilities to apply blockchain technology beyond pure cryptocurrency. For example, the emerging field of decentralized finance revolutionizes great parts of the financial industry (Meyer, Welpe, & Sandner, 2021) with a current market size of $239 billion in 2022, up from $601 million in early 2020 (Amberdata, 2022). Other examples are the arts, gaming and collectibles industries that combined experienced over 21,000% growth with $17.6 billion in sales in 2021 (Quiroz-Gutierrez, 2022) from the NFTs market. NFTs are certificates of ownership, which are stored on a blockchain.

Beyond these megatrends, literature refers to a vast variety of blockchain applications, taking into account different use cases that blockchain offers. The most mentioned blockchain applications are self-sovereign identity, tokenization of assets, fractional ownership, micropayments, smart contracts, and anonymous transactions. Appendix 2 entails an overview of the frequency with which the specific blockchain applications are mentioned by literature. However, we want to address and explain the most frequently discussed blockchain applications briefly here.

In the case of self-sovereign identity, users are able to control their own data and their identity (Toufaily et al., 2021). For example, a blockchain-based ID card and confirmation of residence by the Swiss canton Aargau lets citizens verify their residency without having to disclose information about their identity (Canton of Aargau, 2022). Next, blockchain enables the digital representation of physical assets through tokens, called tokenization, which allows for clear data ownership, reduced fraud and facilitated processing in the blockchain system itself (Abdollahi et al., 2022; Zheng & Boh, 2021; Ziolkowski et al., 2020). Due to decreased cost of verification through disintermediation, property rights can be assigned at a more granular scale (Catalini & Gans, 2016). This way, blockchain enables fractional ownership, as any (illiquid) asset (e.g. a car or house) or a small fraction of it can be traded, exchanged or tracked (Catalini & Gans, 2016). Moreover, through reduced transaction costs with efficient transaction processing, and very small denomination of currency, microtransactions are possible and feasible (Babich & Hilary, 2020; Schlecht et al., 2021). In monetary terms, these are micropayments, such as small on-demand or pay-per-use payments for consumers and creators (Schlecht et al., 2021). For example, the app Fountain lets listeners pay podcast hosts with as little as 1 Satoshi ($1/10^6$ Bitcoin) per minute. Some practitioners and scholars suggest that micropayments are one of the most likely upcoming business model developments (Boston Consulting Group, 2019; Schlecht et al., 2021). Tokens can also be used for financial incentive- and reward programs (Zheng & Boh, 2021). Furthermore, blockchain gave rise to smart contracts. These are digital contracts based on pre-defined terms, which are tamper-proof and self-enforcing through automated execution (Cong & He, 2019; Marikyan, Papagiannidis, Rana, & Ranjan, 2021). Thus, smart contracts ensure accurate value transfers among (pseudo)-anonymous stakeholders in the blockchain network (Marikyan et al., 2021). Irrespective of the area of application, use cases of blockchain leverage the benefits of a tamper-proof information system (Bossler & Kroenung, 2022) that enhances the security and privacy of digital transactions (Marikyan et al., 2022).

3.2. Blockchain technology adoption

Technology adoption describes consumers' behavioral decision to use a technology. Understanding antecedents for consumers' technology adoption is an essential part of information systems research (Blut et al., 2022; Davis, 1989). Other concepts have been developed and applied to explain technology adoption, for example the Theory of Reasoned Action (Fishbein, Ajzen, & Belief, 1975), Technology Acceptance Model (Davis, 1989), Diffusion of Innovation (Rogers, 1962), and UTAUT (Venkatesh et al., 2003). The UTAUT provides a particularly broad picture of user acceptance of technology. Blut et al. (2022) present a revisited UTAUT in their paper and suggest that UTAUT should always consider contextual differences. Even more so, studies on technology adoption should relate to users and include personal characteristics and other context specific predictors. This is in line with the TRI, which indicates that personal motivational factors include optimism and innovativeness while discomfort and insecurity present inhibitors for technology adoption (Parasuraman, 2000; Parasuraman & Colby, 2015). Also, user-oriented technology design is more important than selecting the right user (Blut et al., 2022).

Regarding the adoption of blockchain technology, use cases for organizations or entire industries on the disruptive potential of blockchain technology have been a focus of

blockchain research. Adopting an organizational, managerial perspective (Liang et al., 2021), past research has, for example, looked at business model innovation (Chong et al., 2019; Frizzo-Barker et al., 2020; Weking et al., 2020), its use cases in operations and supply chain (Klöckner, Schmidt, & Wagner, 2022), the private and public sector (Toufaily et al., 2021), the insurance industry (Zhang et al., 2021), as well as opportunities in industry 4.0 (Olsen & Tomlin, 2020), and global shipping (Sarker, Henningsson, Jensen, & Hedman, 2021). But an organizational or industry perspective differs from the consumer's perspective. A consumer's usage intention is a prerequisite for actual usage (Blut et al., 2022; Venkatesh et al., 2003) and should be studied on an individual level, also combining personal predictors, like personal innovativeness, and context specific factors (Blut et al., 2022). Some studies on the adoption of cryptocurrency indicate that knowledge about cryptocurrencies and associated trust could be drivers of cryptocurrency usage (Steinmetz, von Meduna, Ante, & Fiedler, 2021). However, most papers take a general perspective on blockchain technology adoption. Thus, the current state of research lacks an understanding of the usefulness perceptions for specific blockchain applications from a consumer's perspective. Identifying those drivers for perceived usefulness of specific applications would help to better address consumers' motivations and, in turn, influence adoption.

Building on the tenets and findings from technology adoption research in general and blockchain technology adoption studies in particular, we inform our hypotheses on the drivers of usage intention and on perceived usefulness of specific blockchain applications.

4. Research model and hypotheses derivation

4.1. Research model

To better understand the usage intention of blockchain technology, we conflate the abovementioned aspects into two research models (Figure 6 and Figure 7).

Research model I. Technology adoption literature suggests usage intention *increases* in case of higher optimism (H1a), personal innovativeness (H2a), social influence (H5a), trust (H7a), perceived benefits for society (H9a), potential of disruption (H10a), and perceived usefulness across specific applications (H11a). Usage intention *decreases* in case of higher discomfort (H3a), insecurity (H4a), disposition to privacy (H6a), and perceived risk (H8a). To enhance the level of contextualization, we examine interaction effects for age, gender, experience, and cryptocurrency possession.

Research model II. Analogously, application usefulness *increases* with higher optimism (H1b), personal innovativeness (H2b), social influence (H5b), trust (H7b), perceived benefits for society (H9b), and potential of disruption (H10b). Application usefulness *decreases* in case of higher discomfort (H3b), insecurity (H4b), disposition to privacy (H6b) and perceived risk (H8b).

Research model I focuses on usage intention of blockchain technology in general, also searching for moderating effects. We examine the established moderators gender, age, and experience (Venkatesh et al., 2003) and consider contextual differences among consumers by adding possession of cryptocurrency as a moderator (Blut et al., 2022). This is supported by extant literature, as advocating cryptocurrencies is linked to accelerating the pace of blockchain adoption (Catalini & Gans, 2016; Toufaily et al., 2021). Therefore, incorporating cryptocurrency possession is a blockchain specific contextualization measure on the individual level. Instead of articulating distinct hypotheses for all moderating effects, we offer results on those relationships that are observed to be significant.

Taking it a step further, research model II differentiates among the usefulness of six blockchain technology applications, namely: self-sovereign identity, tokenization, fractional ownership, micropayments, smart contracts, and (pseudo-)anonymous transactions, thus providing a more granular view on blockchain technology. Additionally, in line with Blut et al.'s (2022) revised version of UTAUT, we consider a large set of context-aware endogenous mechanisms to study blockchain technology adoption. Table 7 provides an overview of the descriptions of specific blockchain applications.

4.2. Hypotheses derivation

H1; **H2**; **H3**; **H4**. In line with the TRI by Parasuraman and Colby (2015), on the one hand, personal innovativeness and optimism towards new technology are important drivers to predict the technology adoption decision (Blut et al., 2022; Jokisch, Schmidt, Doh, Marquard, & Wahl, 2020; Parasuraman, 2000). On the other hand, discomfort and insecurity regarding new technology hinder technology adoption.

H5. Consumers are influenced by the degree to which important others, such as friends and family, believe a technology should be used (Blut et al., 2022; Venkatesh et al., 2003).

H6. Blockchain technology's transparent nature contrasts with peoples' need for privacy (Raddatz et al., 2021). As transactions in blockchains are transparent and pseudonymous, privacy concerns might arise in the consumer (Rossi et al., 2019).

H7. Literature indicates that consumers' trust in blockchain technology is a key prerequisite to establish relationships and interactions in peer-to-peer markets (Hawlitschek et al., 2016). Trust is established when blockchain technology is perceived as having benevolence, integrity and ability (Hawlitschek et al., 2016). Moreover, prior research suggests that cryptocurrency ownership is driven by trust (Steinmetz et al., 2021).

H8. Risk perception refers to the degree to which consumers have beliefs about potential negative outcomes when using a technology. Therefore, a higher risk perception hinders technology adoption (Koohikamali, Gerhart, & Mousavizadeh, 2015; Pavlou, 2003).

H9. Using blockchain technology can also bring along benefits for society. For instance, economic growth via finan-

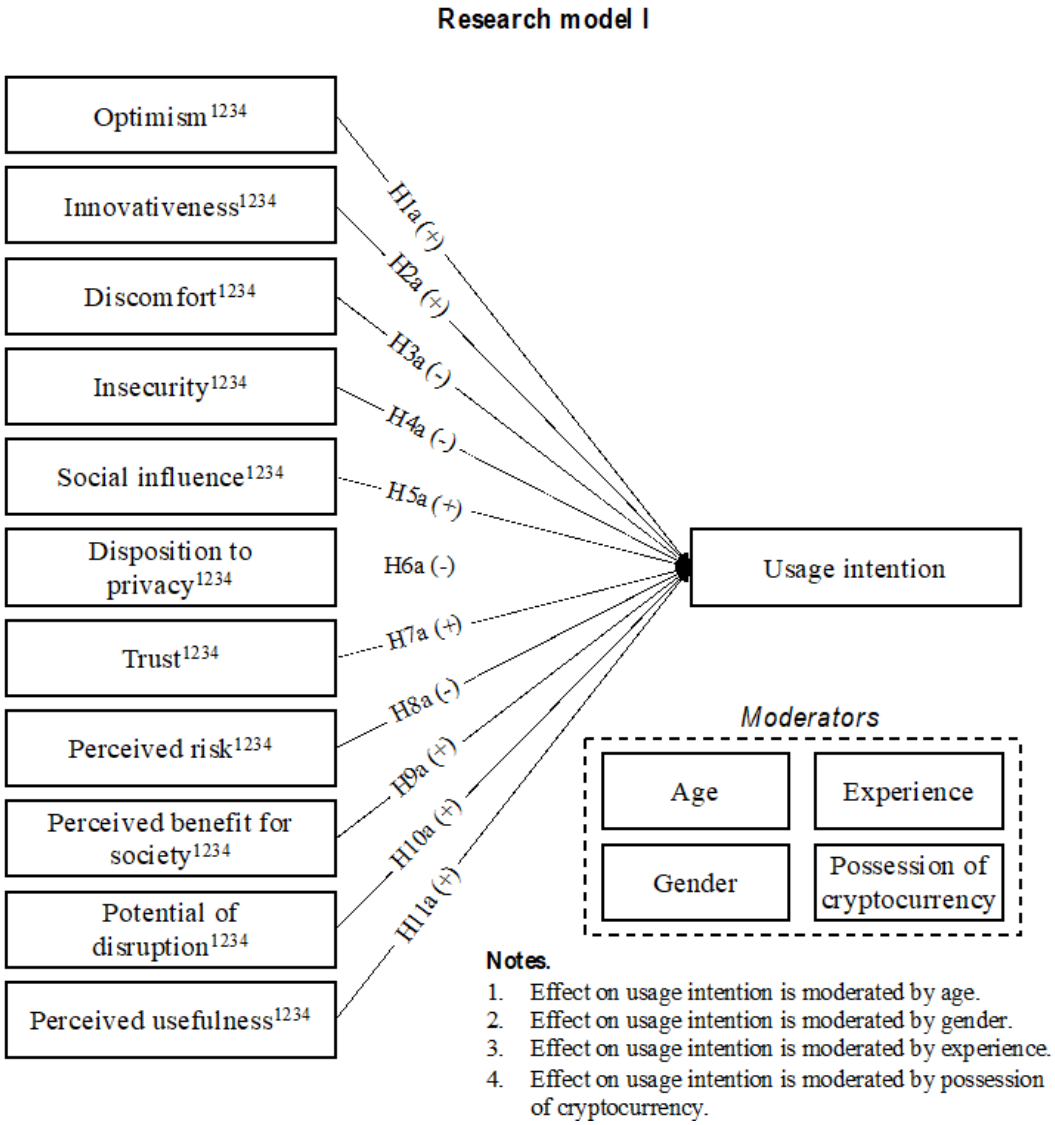


Figure 6: Research model I.

cial and social inclusion (Toufaily et al., 2021). Thus, an expected societal gain may lead to a higher blockchain adoption (Koohikamali et al., 2015).

H10. Literature attributes blockchain technology to be of disruptive nature for business, society and everyday life (Aydiner, 2021; Frizzo-Barker et al., 2020). Hence, consumers' expected efficiency gains will result in an increased adoption.

H11. Originally introduced by Venkatesh et al. (2003) as performance expectancy, its roots are conceptually identical to perceived usefulness (Blut et al., 2022). Consumers are more likely to use transaction technologies such as blockchain if they find them useful. Thus, consumers' expected usefulness of specific blockchain applications drives overall blockchain adoption (Blut et al., 2022; Loh, Lee, Tan, Ooi, & Dwivedi, 2020; Venkatesh et al., 2003).

Beyond the revised UTUAT, we draw upon literature on technology acceptance focused on blockchain technology and also consider literature related to people's technological affinity and possible societal implications. Table 8 grants an overview over the used constructs with context specific definitions.

5. Methodology

5.1. Data analysis

We tested our research model I and II and its associated hypotheses by applying a multiple regression analysis on the German sample population in RStudio 2021.09.1 (view section 2.1 for a summary of the data collection and Table 1 for an overview of the German sample baseline characteristics). With the aim of identifying significant predictors of usage intention (research model I) and application usefulness (research model II) with regards to blockchain technology, it was necessary to maintain comparability of the regression outputs. This comparability of estimates of effects of different variables is a key advantage of path-analytic models

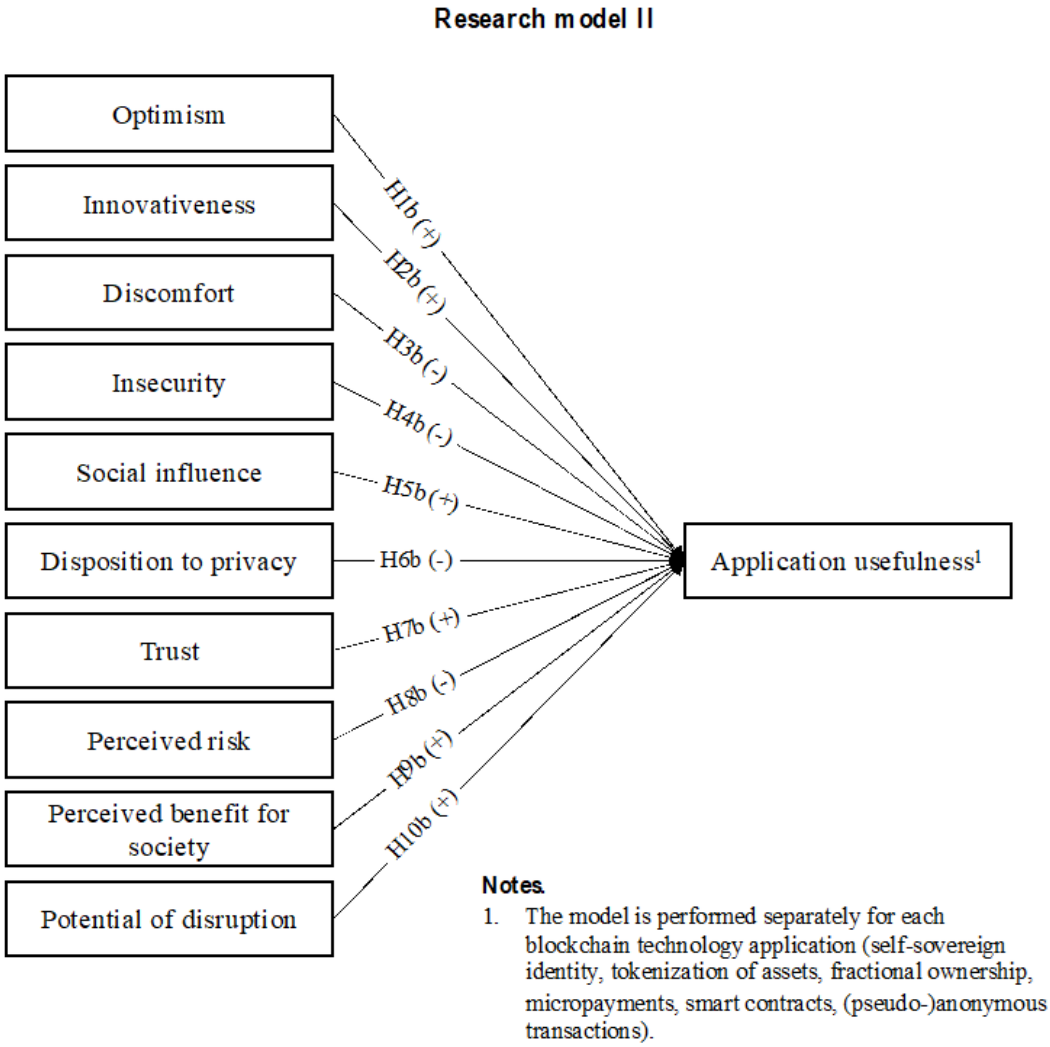


Figure 7: *Research model II.*

such as multiple regression (J. Cohen, Cohen, West, & Aiken, 2003). Therefore, we deliberately chose a multiple regression approach over other commonly used approaches, such as hierarchical regression. Neither did theory constrain an order of predictors before performing the analysis (B. H. Cohen, 2013), nor did our focus lay on assessing the change in predictability that would result from adding further independent variables to the previous included predictors (J. Cohen et al., 2003). Note that multiple regression is equivalently applicable for moderation analysis as is hierarchical regression, as they are mathematically identical and yield the same answer in this respect (Hayes, 2018). As introduced by Venkatesh et al. (2003) in the UTAUT, moderators were applied for tests on usage intention (model I), but not on application usefulness (model II) in our analysis.

5.2. Measures

To ensure content validity, we used validated scales and adapted them to the context of this study. A seven-point Likert scale, ranging from 1 (*strongly disagree*; *very low; not useful at all*) to 7 (*strongly agree; very high; very useful*), was used for the measurement of the items of usage intention, application usefulness, innovativeness, discomfort, insecurity, disposition to privacy, trust, perceived risk, perceived benefits for society, potential of disruption and experience. The arithmetic mean was used to quantify all multi-item constructs.

5.2.1. Dependent variables

This research consists of two separately tested dependent variables, namely usage intention for blockchain technology and specific application usefulness. Usage intention was adapted from UTAUT introduced by Venkatesh et al. (2003) and measured with two items. The first item stated whether the respondent would use blockchain technology applications, the second whether it is very likely that they would use it. To measure the construct of application usefulness, six specific applications were derived from our systematic literature review (Table 7). The survey participants were presented with a short scenario-based description of each application before separately assessing its usefulness. Thus, the

Table 7: Descriptions of specific blockchain applications.

	Application	Description for consumers	Source
I	**Self-sovereign identity**	Details about your identity are digitally stored and you can make selections of it available to others.	Hendershott, Zhang, Zhao, & Zheng, 2021; Toufaily et al., 2021; Zhang et al., 2021
II	**Tokenization of assets**	A real-world item (asset) has a unique, uncopiable, digital representation (token).	Toufaily et al., 2021; Zhang et al., 2021
III	**Fractional ownership**	You can own parts of any real world or digital item or asset.	Kim, 2020; Whitaker & Kräussl, 2020
IV	**Micropayments**	Actions online can trigger micropayments for consumers and creators of content.	Ilk, Guangzhi, Shaokun, & Zhao, 2021-06; Schlecht et al., 2021; Ziolkowski et al., 2020
V	**Smart contracts**	You program a contract digitally and the contract is only fully executed if certain contract details are met. Contracts are not changeable once initiated.	Cong & He, 2019; Frizzo-Barker et al., 2020; Marikyan et al., 2022; Rossi et al., 2019; Schlecht et al., 2021
VI	**Anonymous transactions**	Transactions are possible without having to expose your full identity; only a pseudonym like "8s7dasllsdudmmy8".	Raddatz et al., 2021; Zheng & Boh, 2021; Ziolkowski et al., 2020

research model II was run separately for each application to respectively identify significant predictors.

5.2.2. Independent variables

The items on optimism address whether new technology gives the participant more freedom of mobility and whether new technology makes them more productive.

The items of the innovativeness construct consider firstly whether other people come to the participant for advice on new technologies, secondly whether they are among the first of their friends to acquire new technology, and thirdly whether they keep up with the latest technological developments.

The items on discomfort address whether the respondent could figure out new high-tech products independently and whether they think that technology systems are not designed for use by ordinary people.

The items covering insecurity ask whether the participant believes that people are too dependent on technology, whether too much technology distracts people and whether they do not feel comfortable doing business if the other party is only available online.

Social influence was measured as a single item, inquiring whether the respondent's circle of friends believes that they should use blockchain technology. The scale ranged from 1 (*they would discourage me*) to 10 (*they would encourage me*) but was adjusted to the level of a seven-point Likert scale before analysis.

Items covering disposition to privacy measured participant's sensitivity towards people or organizations handling personal information, the importance of keeping personal information private, and whether the respondent is less concerned about threats to their personal privacy.

The construct of trust is three-dimensional. Items on integrity cover whether the respondent believes that blockchain technology provides reliable information, is honest in dealing with private data, and adheres to principles. Items on benevolence ask about whether the participant thinks that blockchain technology acts in the interest of its users, is not malicious and has no bad intentions. Lastly, items on ability address whether blockchain technology serves its purpose, operates flawlessly and is capable to offer good service.

Perceived risk is quantified using two items, that inquire whether the respondents believe blockchain is risky and whether they feel unsafe using blockchain technology.

Perceived benefit for society was measured by means of two items, examining whether the participant believes that using blockchain technology has advantages for society and whether it has disadvantages.

Potential of disruption was measured using four items, which address whether the respondent thinks that blockchain technology has great potential to disrupt the business world or everyday life, whether it would be as disruptive as the internet or whether it has no disruptive potential at all.

By computing the arithmetic mean of all specific applications usefulness assessments, we measured blockchain's overall perceived usefulness for research model I.

Table 8: Construct variables.

Construct	Context specific definition	Source
Optimism	A consumer's positive view of technology.	Parasuraman (2000)
Innovativeness	A consumer's willingness to try out new technology.	Agarwal & Prasad, 1998; Parasuraman, 2000
Discomfort	A consumer's perceived lack of control over technology and a feeling of being overwhelmed by it.	Parasuraman, 2000
Insecurity	A consumer's distrust of technology and skepticism about its ability to work properly.	Parasuraman, 2000
Social influence	A consumer's perception that others believe they should use blockchain technology.	Venkatesh et al., 2003
Disposition to privacy	A consumer's desire or need for privacy regarding personal information.	Li, 2014
Trust	The believe that blockchain does what they expect from it.	Hawlitschek et al., 2016; Lu, Zhao, & Wang, 2010
Perceived risk	The consumer's beliefs about potential negative outcomes from using blockchain technologies.	Koohikamali et al., 2015
Perceived benefit for society	The consumer's belief of how beneficial blockchain will be for society in general.	Koohikamali et al., 2015
Potential of disruption	The consumer's perception that blockchain technology can fundamentally change businesses or everyday life.	Aydiner, 2021; Frizzo-Barker et al., 2020
Perceived usefulness	The perceived degree to which technology will provide benefits to the consumer across blockchain applications.	c.f. Performance Expectancy Venkatesh et al., 2003
Experience	A consumer's exposure to blockchain technology.	Blut et al., 2022; Venkatesh et al., 2003
Possession of cryptocurrency	A consumer was in possession of cryptocurrency at some point in his or her life.	Steinmetz et al., 2021; Toufaily et al., 2021
Usage intention	The extent to which a person has conscious plans to use blockchain technology.	Venkatesh et al., 2003; Warshaw & Davis, 1985
Application usefulness	The perceived degree to which a specific blockchain application will provide benefits to the consumer.	Venkatesh et al., 2003

5.2.3. Moderator variables

Research model I consists of four moderating variables: age, gender, experience, and possession of cryptocurrency. Consistent with prior research, age was coded as a continuous variable and gender as a 0/1 dummy variable for women and men, respectively (Venkatesh et al., 2003). Experience was operationalized by self-assessed level of knowledge – scale of 1 (*no knowledge*) to 10 (*expert knowledge)* – and the amount of contact to blockchain in the participant's life – professional and private. Knowledge was rescaled before the analysis to the level of a seven-point Likert scale. We applied a 0/1 dummy variable on whether the respondent possess(ed) cryptocurrency or not. The model was run for each moderator respectively.

In sum, the two regression equations in this study are:

1. $$Usage\ intention = b_0 + b_1 Optimism + b_2 Innovativeness + b_3 Discomfort + b_4 Insecurity + b_5 Social\ influence + b_6 Disposition\ to\ privacy + b_7 Trust + b_8 Perceived\ risk + b_9 Perceived\ benefit\ for\ society + b_{10} Potential\ of\ disruption + b_{11} Perceived\ usefulness + b_{12} M + b_{13} Optimism \times M + b_{14} Innovativeness \times M + b_{15} Discomfort \times M + b_{16} Insecurity \times M + b_{17} Social\ influence \times M + b_{18} Disposition\ to\ privacy \times M + b_{19} Trust \times M + b_{20} Perceived\ risk \times M + b_{21} Perceived\ benefit\ for\ society \times M + b_{22} Potential\ of\ disruption \times M + b_{23} Perceived\ usefulness \times M;$$

2. $$Application\ usefulness = b_0 + b_1 Optimism + b_2 Innovativeness + b_3 Discomfort + b_4 Insecurity + b_5 Social\ influence + b_6 Disposition\ to\ privacy + b_7 Trust + b_8 Perceived\ risk + b_9 Perceived\ benefit\ for\ society + b_{10} Potentialof\ disruption,$$

in which M represents the moderating variables age, gender, experience, and possession of cryptocurrency. Appendix 3 provides an overview of the items used to measure the constructs of this study.

5.2.4. Reliability and validity

All variables in multivariate analysis must be assumed to incorporate some degree of measurement error (Hair, 2014). Therefore, it is necessary to assess the degree of measurement error by firstly addressing the reliability and secondly the validity of any measure (Hair, 2014). Construct reliability refers to the degree of consistency between multiple measurements of a variable or set of variables (Hair, 2014). It was measured using Cronbach's a (Cronbach, 1951), the most widely used reliability coefficient (Hair, 2014). As originally introduced by Nunnally (1978), measured variables representing latent constructs should have a coefficients of at least .7 or higher to demonstrate good reliability (Hair, 2014). With the lowest a coefficient at .709 for the construct of perceived benefit for society, good reliability is established. Table 9 shows the a coefficients of the constructs.

Validity is the extent to which a set of measured indicator variables (e.g., items) is associated with their respective underlying factor (e.g., the unobservable construct) (Brown & Moore, 2012; Hair, 2014). To examine validity, both convergent and discriminant validity of the constructs need to be assessed (Hair, 2014). Convergent validity refers to the degree to which items of a specific construct converge or share a high proportion of variance in common (Hair, 2014). For that purpose, we conducted confirmatory factor analysis (CFA) on the measurement model in R to analyze the factor loadings of the items on their respective construct as well as their average variance extracted (AVE) (Hair, 2014). As suggested by prior research, minimum standardized loadings should be at least .5 or higher (Hair, 2014). Therefore, the second item on discomfort (DIS2; See Appendix 3) with a standardized loading of .292 and the first item of experience (EXP1; See Appendix 3) with a standardized loading of .495 were deleted. The resulting CFA reveals that the lowest factor loading is .564, supporting the criteria of convergent validity. Moreover, all measures exceed the recommended AVE minimum of .5 (Parasuraman & Colby, 2015). Thus, convergent validity of the model is confirmed. See Table 9 for all factor loadings and AVEs based on CFA.

Discriminant validity refers to the extent to which constructs are truly distinct to another, both in terms of their correlations and whether the items represent only their associated construct (Hair, 2014). Due to the limitations of examining discriminant validity based on traditional approaches, like by assessing the Fornell-Larcker criterion or cross-loadings, we instead used the heterotrait-monotrait (HTMT) ratio of correlations (Henseler, Ringle, & Sarstedt, 2015). The HTMT ratio of correlation measures the degree of similarity between constructs (Henseler et al., 2015; Raddatz et al., 2021). Due to potential difficulties in empirically distinguishing constructs in technology acceptance models, HTMT ratios below .9 indicate discriminant validity (Henseler et al., 2015). All HTMT values are below .9, except from a ratio of .934 between the constructs perceived risk and perceived benefit for society. Table 10 provides an overview of the HTMT ratios.

To rule out any multicollinearity issues arising from this result, an analysis of the variance inflation factors (VIF) shows that all non-moderated independent variables are below the recommended threshold of 5 (Hair, 2014). As McClelland, Irwin, Disatnik, and Sivan (2017) suggest, multicollinearity is not a concern for moderator variables. See Table 9 for the results on the VIFs of the constructs.

Moreover, we used CFA to assess the overall measurement model fit to examine whether a high correlation estimate undermines the discriminant validity and unidimensionality of the constructs (Rönkkö & Cho, 2022). The results on the χ^2 index, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) (Hair, 2014) indicate that the measurement model fits the data well

(χ^2 index = 3.383, CFI = .931, TLI = .920, RMSEA = .055 and SRMR = .044). Therefore, acceptable discriminant validity is confirmed. See Table 11 for an overview of the model fit statistics and their recommended values (Brown, 2015; Hair, 2014; Hu & Bentler, 1999).

Note that the reliability and validity of single-item constructs cannot be computed (Hair, 2014), which is why discomfort (a single-item construct after validity testing), social influence, age, gender and possession of cryptocurrency were omitted from reliability and validity criteria. VIFs of discomfort and social influence can be found in the notes of Table 9.

In sum, the items and constructs of the measurement model demonstrate reliability, convergent and discriminant validity as well as a good overall model fit. Thus, the measures can be used confidently for statistical analysis.

6. Results

6.1. Descriptive results

Before performing multiple regression analysis, the means, standard deviations, and correlation coefficients of all variables were derived (Table 12).

Out of the TRI-based constructs, optimism and insecurity show mean values above their scale's center of 4 (4.30 and 4.83), whereas discomfort and innovativeness are below (3.53 and 3.13). On the same scale, the mean value of social influence is 3.30, for disposition to privacy it is 4.60, for trust it is 4.29, for perceived risk it is 4.43, for perceived benefit for society it is 4.13, for potential of disruption it is 4.21 and for experience it is 1.93. Perceived usefulness has a mean of 4.10 and usage intention a mean of 3.22. Implemented as a continuous variable, age has a mean of 48.79. The two dummy variables gender and possession of cryptocurrency demonstrate mean values of .49 and .12, respectively.

A correlation analysis of the variables yields significant correlation coefficients at the 5% level for 95% (114) of all 120 correlation coefficients. 91% (109) of correlation coefficients are significant at the 1% level. 87% (104) of the coefficients remain below .5. However, a correlation analysis does not provide the level of statistical rigor to test the hypothesized relationships in this paper. Thus, the results of the (moderated) multiple regression can be found in the following section.

6.2. Effects on usage intention

For research model I, five regression models were performed on usage intention. Table 13 provides the results of the tests of research model I. As the moderated regressions significantly increase the explained variance compared to the unmoderated model ($DR^2_A = .021$, $p < .01$; $DR^2_B = .010$, $p < .05$; $DR^2_C = .021$, $p < .01$; $DR^2_D = .016$, $p < .01$), the results of the unmoderated model are negligible (Hair, 2014).

Inconsistent with the hypothesis that optimism has a positive effect on usage intention, optimism displays no significant (n.s.) effect (b_{A-D}, n.s.). Thus, H1a is not supported. As predicted, innovativeness shows a statistically significant positive effect across all models ($b_A = .20$, $p < .05$; $b_B = .13$, $p < .01$; $b_C = .15$, $p < .05$; $b_D = .12$, $p < .01$). Therefore, H2a is confirmed.

In line with the hypothesis, discomfort shows a pattern of negative effects on usage intention in three out of four models ($b_A = .04$, n.s.; $b_B = -.08$, $p < .05$; $b_C = -.13$, $p < .05$; $b_D = -.06$, $p < .05$). Thus, H3a is confirmed in the models including gender, experience, or cryptocurrency possession. Inconsistent with the hypothesized relationship, no significant effect is observed for insecurity (b_{A-D}, n.s.). Hence, H4a is not confirmed. As hypothesized, social influence has a positive effect on usage intention, which is significant in three out of four models ($b_A = .07$, n.s.; $b_B = .15$, $p < .01$; $b_C = .10$, $p < .01$; $b_D = .10$, $p < .01$). Therefore, H5a is confirmed in the models including gender, experience, or cryptocurrency possession. No significant relationship is found for disposition to privacy (b_{A-D}, n.s.). Consequently, H6a is not supported.

As predicted, there is a consistent pattern that trust positively affects usage intention ($b_A = .81$, $p < .01$; $b_B = .19$, $p < .01$; $b_C = .02$, n.s.; $b_D = .17$, $p < .01$). As the interaction effect of trust and experience is significant in model C ($b_C = .11$, $p < .01$), the positive effect of trust is observed in model C as well, although the simple unmoderated effect is not significant. Thus, H7a is supported. However, the moderators age, experience and possession of cryptocurrency affect the relationship between trust and usage intention significantly. Specifically, the positive effect of trust on usage intention decreases with an increase in age ($b_A = -.01$, $p < .01$), it increases with an increase in experience ($b_C = .11$, $p < .01$) and it increases with the possession of cryptocurrency ($b_D = .33$, $p < .05$).

In line with the prediction, perceived risk affects usage intention negatively, which is significant in three out of four models ($b_A = -.07$, n.s.; $b_B = -.17$, $p < .01$; $b_C = -.31$, $p < .01$; $b_D = -.28$, $p < .01$). Therefore, H8a is confirmed in the models including gender, experience, or cryptocurrency possession. Moreover, the moderators gender and possession of cryptocurrency significantly affect this relationship. Specifically, the negative effect of perceived risk on usage intention increases for males ($b_B = -.19$, $p < .01$) and it decreases with the possession of cryptocurrency ($b_D = .32$, $p < .01$).

Inconsistent with the hypothesis, no significant effect is observed for perceived benefit for society (b_{A-D}, n.s.). Hence, H9a is not supported. As hypothesized, potential of disruption has a positive effect on usage intention, which is significant in three out of four models ($b_A = .31$, $p < .01$; $b_B = .13$, $p < .05$; $b_C = .12$, n.s.; $b_D = .18$, $p < .01$). Thus, H10a is supported in the models including age, gender, or cryptocurrency possession. There is a consistent pattern that perceived usefulness has a positive effect on usage intention, in line with the prediction ($b_A = .24$, $p < .05$; $b_B = .23$, $p < .01$; $b_C = .33$, $p < .01$.; $b_D = .27$, $p < .01$). Consequently, H11a is confirmed. Furthermore, experience significantly moderates the relationship between perceived usefulness and us-

Table 9: CFA results.

Construct	Item[a]	Loading > .5	CR α > .7	AVE > .5	VIF[b] < 5
Optimism			.837	.720	2.063
	OPT1	.849			
	OPT2	.848			
Innovativeness			.864	.682	2.104
	INN1	.841			
	INN2	.847			
	INN3	.791			
Insecurity			.758	.520	1.271
	INS1	.634			
	INS2	.772			
	INS3	.740			
Disposition to privacy			.753	.521	1.137
	DTP1	.756			
	DTP2	.828			
	DTP3	.564			
Trust			.949	.680	2.771
	TIN1	.854			
	TIN2	.811			
	TIN3	.841			
	TBE1	.830			
	TBE2	.837			
	TBE3	.842			
	TAB1	.804			
	TAB2	.785			
	TAB3	.805			
Perceived risk			.773	.639	2.085
	RIS1	.763			
	RIS2	.829			
Perceived benefit for society			.709	.566	3.362
	BSO1	.838			
	BSO2	.655			
Potential of disruption			.869	.641	2.179
	PDI1	.884			
	PDI2	.920			
	PDI3	.800			
	PDI4	.579			
Perceived usefulness			.884	.559	2.103
	USF1	.774			
	USF2	.746			
	USF3	.816			
	USF4	.755			
	USF5	.667			
	USF6	.739			
Usage intention			.960	.924	-
	UIN1	.953			
	UIN2	.969			
Experience			.724	.604	-
	EXP2	.731			
	EXP3	.803			

Note: CFA was applied using the "lavaan" package in R, which reduced N to 787 for this purpose. AVE = Average Variance Extracted; CR a = Cronbach's a; VIF = Variance Inflation Factor. [a]List of all corresponding items can be found in the Appendix 3 [b]VIF of discomfort = 1.657; VIF of social influence = 1.240.

Table 10: HTMT ratios.

	OPT	INN	INS	DTP	TRU	RIS	BSO	PDI	USF	UIN	EXP
OPT	1										
INN	.709	1									
INS	.407	.360	1								
DTP	.219	.128	.363	1							
TRU	.551	.370	.190	.214	1						
RIS	.387	.391	.347	.249	.643	1					
BSO	.547	.372	.330	.273	.891	.934	1				
PDI	.492	.291	.122	.134	.732	.556	.841	1			
USF	.535	.351	.195	.186	.698	.547	.794	.713	1		
UIN	.524	.515	.257	.184	.693	.719	.779	.661	.700	1	
EXP	.378	.656	.211	.091	.387	.474	.408	.362	.396	.596	1

Note: N = 847. OPT = Optimism; INN = Innovativeness; INS = Insecurity; DTP = Disposition to privacy; TRU = Trust; RIS = Perceived risk; BSO = Perceived benefits for society; PDI = Potential of disruption; USF = Perceived usefulness; UIN = Usage intention; EXP = Experience.

Table 11: CFA model fit statistics.

Goodness-of-fit statistic	Recommended value	Computed value
χ^2 (Chi-square)	-	2063.595
Degrees of freedom	-	610
p-value of χ^2	-	.000
χ^2 index (χ^2 / degrees of freedom)	< 5 (Hair, 2014)	3.383
Comparative Fit Index (CFI)	$\geq .9$ (Hair, 2014)	.931
Tucker-Lewis Index (TLI)	$\geq .9$ (Hair, 2014)	.920
Root Mean Square Error of Approximation (RMSEA)	$\leq .06$ (Hu & Bentler, 1999)	.055
Standardized Root Mean Square Residual (SRMR)	$\leq .08$ (Hu & Bentler, 1999)	.044

Note: CFA was applied using the "lavaan" package in R, which reduced N to 787 for this purpose.

age intention. The positive effect of perceived usefulness on usage intention decreases with an increase in experience ($b_C = -.05$, $p < .05$).

6.3. Effects on application usefulness

For research model II, six regression models were performed on application usefulness. See Table 14 for the results of the tests of research model II.

As predicted, optimism is observed to have a positive effect on the application usefulness of blockchain technology. A significant effect is found for five out of six applications ($b_{TOA} = .16$, $p < .01$; $b_{FOW} = .13$, $p < .05$; $b_{SSI} = .21$, $p < .01$.; $b_{SCO} = .13$, $p < .05$; $b_{MPY} = .06$, n.s.; $b_{ATR} = .14$, $p < .01$). Therefore, H1b is supported for every application, except for micropayments. Inconsistent with the hypotheses, no significant effect is observed for the other TRI-based constructs of innovativeness ($b_{TOA-ATR}$, n.s.), discomfort ($b_{TOA-ATR}$, n.s.) and insecurity ($b_{TOA-ATR}$, n.s.). Thus, H2b, H3b and H4b are not confirmed. The predicted positive effect of social influence on application usefulness can only be observed for the applications tokenization of assets and fractional ownership ($b_{TOA} = .06$, $p < .05$; $b_{FOW} = .08$, $p < .01$). Consequently, H5b is confirmed for tokenization and fractional ownership applications. Disposition to privacy shows a negative effect on the application usefulness of self-sovereign identity and smart contracts, which is in line with the hypothesis ($b_{SSI} = -.08$, $p < .05$; $b_{SCO} = -.15$, $p < .01$). Hence, H6b is confirmed for self-sovereign identity and smart contract applications. As hypothesized, trust has a positive effect on application usefulness ($b_{TOA} = .26$, $p < .01$; $b_{FOW} = .21$, $p < .01$; $b_{SSI} = .28$, $p < .01$.; $b_{SCO} = .20$, $p < .01$; $b_{MPY} = .31$, $p < .01$; $b_{ATR} = .22$, $p < .01$). Therefore, H7b is supported for every application. Inconsistent with the predicted relationship, no significant effect is observed for perceived risk ($b_{TOA-ATR}$, n.s.). Thus, H8b is not confirmed. In line with the prediction, perceived benefit for society positively affects application usefulness. A significant effect is observed for five out of six applications ($b_{TOA} = .35$, $p < .01$; $b_{FOW} = .19$, $p < .01$; $b_{SSI} = .34$, $p < .01$.; $b_{SCO} = .35$, $p < .01$; $b_{MPY} = .12$, n.s.; $b_{ATR} = .28$, $p < .01$). Consequently, H9b is supported for every application, except for micropayments. As predicted, there is a

Table 12: Descriptive statistics.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. OPT	4.30	1.48															
2. INN	3.13	1.57	.60**														
3. DIS	3.53	1.67	-.48**	-.60**													
4. INS	4.83	1.28	-.32**	-.27**	.11**												
5. SOC	3.30	1.92	.31**	.33**	-.25**	-.16**											
6. DTP	4.60	1.33	-.17**	-.05	.06	.28**	-.09*										
7. TRU	4.29	1.31	.46**	.32**	-.27**	-.15**	.33**	-.19**									
8. RIS	4.43	1.52	-.29**	-.31**	.23**	.26**	-.32**	.20**	-.55**								
9. BSO	4.13	1.34	.41**	.28**	-.20**	-.23**	.32**	-.21**	.73**	-.69**							
10. PDI	4.21	1.44	.40**	.25**	-.18**	-.10**	.28**	-.11**	.66**	-.45**	.65**						
11. USF	4.10	1.53	.45**	.31**	-.22**	-.15**	.32**	-.17**	.64**	-.46**	.63**	.62**					
12. UIN	3.22	1.69	.44**	.44**	-.34**	-.20**	.44**	-.16**	.66**	-.61**	.64**	.60**	.64**				
13. AGE	48.79	17.07	-.16**	-.17**	.24**	.00	-.24**	.13**	-.12**	.19**	-.17**	-.07*	-.19**	-.26**			
14. GEN	.49	.50	.11**	.30**	-.22**	-.07*	.08*	.01	.11**	-.12**	.10**	.07*	.12**	.17**	.15**		
15. EXP	1.93	1.21	.29**	.52**	-.34**	-.14**	.45**	-.05	.32**	-.35**	.30**	.29**	.32**	.49**	-.17**	.27**	
16. POC	.12	.32	.15**	.26**	-.19**	-.06	.24**	-.10**	.23**	-.29**	.20**	.21**	.22**	.36**	-.16**	.17**	.56**

Note: N = 847. M = Mean; SD = Standard deviation; OPT = Optimism; INN = Innovativeness; DIS = Discomfort; INS = Insecurity; SOC = Social influence; DTP = Disposition to privacy; TRU = Trust; RIS = Perceived risk; BSO = Perceived benefits for society; PDI = Potential of disruption; USF = Perceived usefulness; UIN = Usage intention; AGE = Age; GEN = Gender; EXP = Experience; POC = Possession of cryptocurrency. * $p < .05$. ** $p < .01$.

Table 13: Regression results of research model I: Usage intention.

Dependent variable: *Usage intention*	Unmoderated regression			Moderated regression (A) Age			(B) Gender			(C) Experience			(D) POC		
	β	*t*	*p*	*β*	*t*	*p*	*β*	*t*	*p*	*β*	*t*	*p*	*β*	*t*	*p*
Direct effects															
Intercept	.98*	2.57	.01	-.62	-.51	.61	.45	.82	.41	1.91**	2.81	.01	1.21**	2.99	.00
Optimism	-.03	-.75	.45	-.11	-1.01	.31	.03	.60	.55	.00	.06	.95	-.01	-.20	.84
Innovativeness	.15**	4.61	.00	.20*	2.08	.04	.13**	2.82	.00	.15*	2.27	.02	.12**	3.45	.00
Discomfort	-.05	-1.91	.06	.04	.54	.59	-.08*	-2.16	.03	-.13*	-2.36	.02	-.06*	-2.01	.04
Insecurity	.00	.09	.93	-.03	-.26	.79	.00	.02	.99	-.02	-.37	.71	-.01	-.25	.80
Social influence	.11**	5.62	.00	.07	1.16	.25	.15**	5.25	.00	.10**	2.77	.01	.10**	4.65	.00
Disposition to privacy	.00	-.08	.93	-.15	-1.72	.09	-.01	-.31	.75	-.03	-.64	.52	.01	.24	.81
Trust	.21**	4.84	.00	.81**	5.68	.00	.19**	3.05	.00	.02	.23	.82	.17**	3.74	.00
Perc. risk	-.27**	-8.00	.00	-.07	-.67	.50	-.17**	-3.51	.00	-.31**	-5.13	.00	-.28**	-7.82	.00
Perc. benefit for society	.03	.55	.58	-.23	-1.48	.14	.12	1.75	.08	.05	.55	.58	.02	.37	.71
Potential of disruption	.18**	5.12	.00	.31**	2.68	.01	.13*	2.44	.01	.12	1.81	.07	.18**	4.80	.00
Perc. usefulness	.26**	7.81	.00	.24*	2.16	.03	.23**	4.85	.00	.33**	5.80	.00	.27**	7.72	.00
Age				.03	1.34	.18									
Gender							.87	1.14	.25						
Experience										-.49	-1.60	.11			
Possession of cryptocurrency													-2.30	-1.96	.05
Moderation effects															
M × Optimism				.00	.74	.46	-.12	-1.79	.07	.00	-.14	.89	-.06	-.56	.58
M × Innovativeness				.00	-.57	.57	.06	.82	.41	-.03	-.98	.33	-.02	-.24	.81
M × Discomfort				.00	-1.06	.29	.07	1.36	.17	.05	1.92	.05	.07	.72	.47
M × Insecurity				.00	.17	.87	.02	.35	.73	.01	.54	.59	.02	.17	.87
M × Social influence				.00	.49	.62	-.07	-1.78	.08	-.01	-.55	.59	-.01	-.22	.82
M × Disposition to privacy				.00	1.92	.06	.02	.38	.71	.02	.89	.37	.02	.27	.78
M × Trust				-.01**	-4.29	.00	.05	.58	.56	.11**	2.91	.00	.33*	2.31	.02
M × Perc. risk				.00	-1.90	.06	-.19**	-2.80	.01	.04	1.41	.16	.32**	2.86	.00
M × Perc. benefit for society				.01	1.89	.06	-.19	-1.95	.05	.00	-.13	.90	.22	1.54	.12
M × Potential of disruption				.00	-.99	.32	.11	1.49	.14	.02	.79	.43	.01	.06	.95
M × Perc. usefulness				.00	-.07	.94	.06	.97	.33	-.05*	-1.98	.05	-.18	-1.70	.09
R^2	.644**			.664**			.653**			.664**			.660**		
$\Delta R^2$to unmoderated model				.021**			.010*			.021**			.016**		

Note: N = 847. M = moderator variable, which is a generic representative for the respective moderator of the moderated model (A-D); b = unstandardized regression weight; *t* = t-value. *p* = p-value, POC = Possession of cryptocurrency; *$p < .05$. **$p < .01$.

consistent pattern that potential of disruption has a positive effect. ($b_{TOA} = .23$, $p < .01$; $b_{FOW} = .27$, $p < .01$; $b_{SSI} = .26$, $p < .01$.; $b_{SCO} = .26$, $p < .01$; $b_{MPY} = .30$, $p < .01$; $b_{ATR} = .25$, $p < .01$). Hence, H10b is confirmed for every application. Moreover, the model for every application displays a significant R^2 ($R^2_{TOA} = .399$, $p < .01$; $R^2_{FOW} = .324$, $p < .01$; $R^2_{SSI} = .408$, $p < .01$.; $R^2_{SCO} = .332$, $p < .01$; $R^2_{MPY} = .268$, $p < .01$; $R^2_{ATR} = .317$, $p < .01$). Table 15 provides an overview of the supported and not supported hypotheses investigated in this paper.

To add value to the statistical analysis of application usefulness, we investigated descriptively which specific blockchain applications were considered most useful. The results reveal that self-sovereign identity applications are currently considered most useful (53% of respondents answered between 5 (*somewhat useful*) and 7 (*very useful*) on the Likert scale), followed by tokenization of assets (52%), anonymous transactions (47%), smart contracts (44%), micropayments (44%) and fractional ownership (36%). See Figure 8 for an overview of consumers' usefulness assessments of the specific blockchain applications. See Appendix 4 for an overview of the application usefulness assessments of the British sample population.

Table 14: Regression results of research model II: Application usefulness.

Dependent var.:	Applications																	
Application	TOA			FOW			SSI			SCO			MPY			ATR		
usefulness	β	t	p	β	t	p	β	t	p	β	t	p	β	t	p	β	t	p
Intercept	-.50	-.94	.35	-.30	-.51	.61	-.10	-.17	.86	.41	.69	.49	.39	.57	.57	-.03	-.06	.95
Optimism	.16**	3.44	.00	.13*	2.46	.01	.21**	4.15	.00	.13*	2.37	.02	.06	1.04	.30	.14**	2.69	.01
Innovativeness	.02	.40	.69	.09	1.84	.07	-.02	-.40	.69	-.02	-.34	.74	.07	1.23	.22	.02	.43	.67
Discomfort	.02	.65	.51	.07	1.67	.10	.01	.13	.89	-.02	-.37	.71	.01	.19	.85	.02	.55	.58
Insecurity	.04	1.02	.31	.03	.55	.58	.04	.80	.42	.08	1.65	.10	-.01	-.23	.82	.03	.64	.53
Social influence	.06*	2.13	.03	.08**	2.58	.01	.00	.05	.96	.03	.89	.37	.07	1.88	.06	.03	.85	.40
Disp. to privacy	-.02	-.51	.61	.00	.01	.99	-.08*	-1.98	.05	-.15**	-3.38	.00	-.02	-.40	.69	.06	1.37	.17
Trust	.26**	4.32	.00	.21**	3.12	.00	.28**	4.40	.00	.20**	2.89	.00	.31**	3.91	.00	.22**	3.31	.00
Perc. risk	.04	.95	.34	-.06	-1.26	.21	.03	.57	.57	.01	.25	.80	-.05	-.84	.40	-.04	-.78	.44
Perc. ben. for soc.	.35**	5.26	.00	.19**	2.68	.01	.34**	5.01	.00	.35**	4.71	.00	.12	1.45	.15	.28**	3.82	.00
Pot. of disruption	.23**	4.77	.00	.27**	5.02	.00	.26**	5.21	.00	.26**	4.75	.00	.30**	4.82	.00	.25**	4.62	.00
R^2	.399**			.324**			0.408**			.332**			.268**			.317**		

Note: N = 847. TOA = Tokenization of Assets; FOW = Fractional Ownership; SSI = Self-Sovereign Identity; SCO = Smart Contracts; MPY = Micropayments; ATR = Anonymous Transactions; b = unstandardized regression weight. t = t-value. p = p-value. * $p < .05$. ** $p < .01$.

Table 15: Summary of results of the hypothesized effects.

Hypothesis	Effect on usage intention	Supported $p < .05$	Hypothesis	Effect on application usefulness	Supported $p < .05$
H1a	Optimism (+)	**Not supported**	H1b	Optimism (+)	**Supported**[3]
H2a	Innovativeness (+)	**Supported**	H2b	Innovativeness (+)	**Not supported**
H3a	Discomfort (-)	**Supported**[1]	H3b	Discomfort (-)	**Not supported**
H4a	Insecurity (-)	**Not supported**	H4b	Insecurity (-)	**Not supported**
H5a	Social influence (+)	**Supported**[1]	H5b	Social influence (+)	**Supported**[4]
H6a	Disposition to privacy (-)	**Not supported**	H6b	Disposition to privacy (-)	**Supported**[5]
H7a	Trust (+)	**Supported**	H7b	Trust (+)	**Supported**
H8a	Perceived risk (-)	**Supported**[1]	H8b	Perceived risk (-)	**Not supported**
H9a	Perceived benefit for society (+)	**Not supported**	H9b	Perceived benefit for society (+)	**Supported**[3]
H10a	Potential of disruption (+)	**Supported**[2]	H10b	Potential of disruption (+)	**Supported**
H11a	Perceived usefulness (+)	**Supported**			

Note:[1] Effect confirmed in the models including gender, experience, or possession of cryptocurrency. [2] Effect confirmed in the models including age, gender, or possession of cryptocurrency. [3] Effect confirmed for every specific application, except micropayments. [4] Effect confirmed for tokenization and fractional ownership applications. [5] Effect confirmed for self-sovereign identity and smart contract applications.

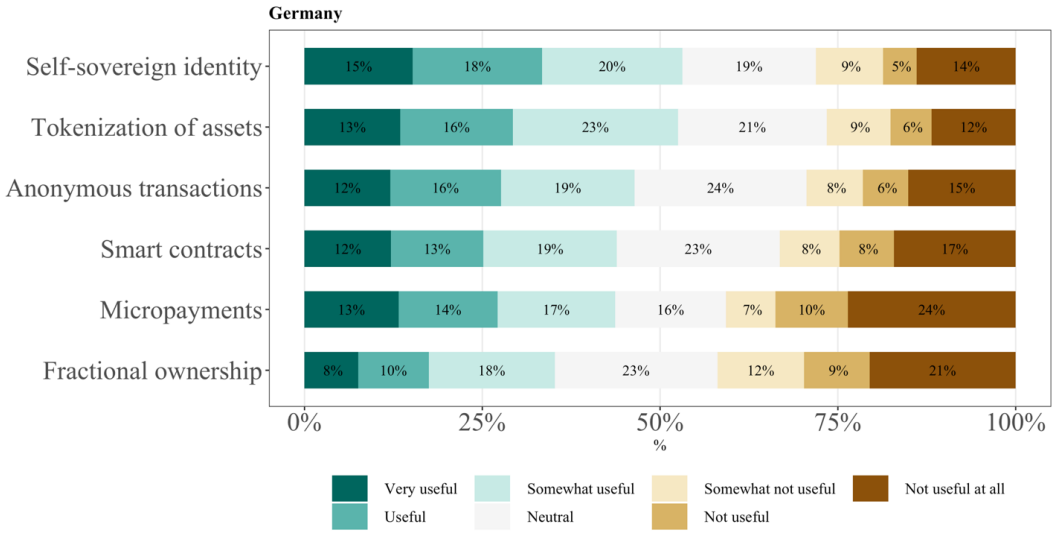


Figure 8: Consumers' usefulness assessment of specific blockchain applications.

Note: N = 847.

7. Discussion

Blockchain technology research has mainly been focused on general usage intention, mostly examined the organizational perspective, and lacked a differentiated view at specific blockchain applications from the viewpoint of the consumer. As we show in this paper, contextual factors influence the relationships of drivers for usage intention. Furthermore, consumers consider certain blockchain applications to be more useful than others. This indicates that blockchain adoption research should be more granular and differentiate between applications and contexts.

Trust and consumers' perceived usefulness are found to be strong, positive drivers of usage intention. Our findings indicate that consumers, who recognize blockchain's inherent integrity, benevolence, and ability, show trust towards the technology that consequently increases their usage intention. This is in line with existing literature on consumer-to-consumer markets, which indicates that trust towards peers and products increases consuming and purchasing intentions (Hawlitschek et al., 2016; Lu et al., 2010).

The result of perceived usefulness is consistent with prior UTAUT studies by Blut et al. (2022) and Venkatesh et al. (2003), who confirmed positive effects on behavioral intention. Therefore, if blockchain applications such as self-sovereign identity or tokenization of assets are designed to be of higher usefulness to the consumer, the consumers' usage intention increases.

In contrast to the privacy value proposition of blockchain technology, consumers' usage intention is not driven by their disposition to privacy. Although previous studies have emphasized the trade-off between risk and benefits for adoption decisions (Marikyan et al., 2022), consumer's beliefs of keeping personal information private do not seem to play a significant role – at least not at the current stage of blockchain adoption. This surprising relationship is in line with Raddatz et al. (2021), who observed no influence of privacy concerns of consumers on their perceived benefits from using blockchain technology. Possible reasons might be that consumers have not yet fully understood the decentralized and transparent characteristics of blockchain or (pseudo-)anonymity fulfills their need for privacy.

Consumers' perceived risk has a strong negative effect on blockchain usage intention. Concerns on system failure, security, reliability or other personal, psychological or financial risks should be minimized to boost adoption (Blut & Wang, 2020). Explorers might even face higher innovation failure risks than Hesitators or Avoiders (Abdollahi et al., 2022).

Social influence shows a weak, but positive effect on blockchain usage intention in the models including gender, experience, or possession of cryptocurrency. This finding confirms prior research (Liang et al., 2021; Venkatesh et al., 2003) and shows that blockchain usage intention is influenced by the people surrounding the consumer. Social influence is a particularly significant factor in the early adoption phase of a new technology, but might become insignificant over time (Liang et al., 2021).

Although prior studies supported a positive relationship between the overall perception of benefits and the attitude of consumers (Koohikamali et al., 2015), narrowing these benefits down to societal benefits shows no influence. This might be because societal benefits such as new economic opportunities, acceleration of peer-to-peer economies, or refined citizen-government interactions (Toufaily et al., 2021) take a long time to be realized and experienced by the consumer.

Consumers' beliefs about the potential of disruption of blockchain show a positive effect in the models including age, gender, or possession of cryptocurrency. This indicates that consumers, who see some disruptive potential of blockchain, have a higher usage intention. Primed by many advocates as

being a "disruptive innovation", this label does not seem to go unnoticed by consumers (Frizzo-Barker et al., 2020).

Consumers' innovativeness has a positive effect on blockchain usage intention. This is in line with Blut and Wang (2020), who observed strong positive, indirect effects of motivators on usage behavior. Moreover, our results are consistent with Blut et al. (2022), who showed a strong association of personal innovativeness with actual usage. Our results reveal that consumers, who are technology pioneers and thought leaders show a higher intention to use blockchain technology. Therefore, addressing lead-users and the Explorer segment of the population is critical when aiming to foster widespread blockchain adoption.

Interestingly, inhibitory forces (specifically discomfort) of the TRI show a weaker effect on usage intention than the motivational forces (specifically innovativeness). This is consistent with findings from Blut and Wang (2020) on technology usage. Thus, this study supports existing literature by indicating that consumers do not feel in control of blockchain technology and are somewhat overwhelmed by it (Marikyan et al., 2022). Alleviating their discomfort and fostering their understanding of blockchain technology is crucial to enhance blockchain adoption.

7.1. Theoretical contributions

This paper makes five contributions to blockchain adoption research. First, this is one of the first papers to identify and investigate the drivers of blockchain usage intention from the perspective of the consumer by combining streams of technology adoption literature. Our results refine current UTAUT-, TRI-, and blockchain specific theory and reveal which predictors are relevant in the context of blockchain adoption. Second, our study shows the relevance of including individual characteristics and context specific moderators, such as possession of cryptocurrency. Past research has commonly focused on the main effect of predictors, specifically for UTAUT, while neglecting contextual differences (Blut et al., 2022). Third, as called for by Rossi et al. (2019), we systematically identify specific blockchain applications for future research to build upon. Our findings reveal which specific applications might be most promising from the perspective of the consumer. Fourth, distinguishing between general usage intention and specific application usefulness enables us to provide an indication on which predictors are more important for which specific application. Lastly, we provide a field report on the perception of blockchain technology by consumers in Germany and the UK as well as a cluster analysis based on the technology readiness of the German and British population. This provides research with a status quo and allows for contextualization in technology adoption research.

7.2. Practical contributions

Based on our results, we put forward guiding principles for business managers and blockchain organizations to influence the adoption of blockchain technology. To boost the general intention to use blockchain, managers need to appeal to a consumer group that, on the one hand, contains a) innovative people, b) who recognize the usefulness of the specific application, c) who are influenced to a certain degree by their social environment, d) who show higher levels of trust in blockchain technology and e) credit blockchain some disruptive potential. Based on technology readiness, Explorers and Pioneers are most likely to fit this description.

Managers should utilize the public characteristic of blockchain and enable employees to experiment with it. Blockchain is easily accessible, even though its user interface is still in its infancy. Yet, managers need to alleviate perceived risks and concerns of discomfort of consumers. This could be achieved by e.g., designing user-oriented front ends of applications, providing a proper onboarding process, or encouraging hands-on experiences by giving out free product trials.

Organizations need to take into account age, gender, experience and cryptocurrency possession. First, managers should appeal to younger consumers by communicating technological features that convey benevolence, ability, and integrity of blockchain. For example, that the Bitcoin blockchain operates flawlessly since inception. Second, although men are more prevalent in the Explorer and Pioneer segment, their relationship between perceived risk and usage intention is more sensitive than it is among women. This indicates that men have more knowledge of blockchain technology, consider more risk factors and are more aware of the downfalls of blockchain technology, thus their usage intention is reduced. Therefore, managers should bear in mind that even though young men appear to be more inclined to use blockchain, it is critical to also reduce their perceived risk. Third, managers should aim to increase consumers' experience levels with blockchain technology. As consumers gain more knowledge about blockchain and their exposure to blockchain increases, trust seems to become more important to the consumer than their perceived usefulness of the application. Therefore, managers should aim at increasing knowledge of consumers on blockchain and getting more consumers into contact with blockchain through e.g., free product versions or social media marketing campaigns with free training documents. Fourth, mangers should give potential customers cryptocurrency to incentivize blockchain adoption. Hands-on experience reduces the impact of consumers' perceived risks when using blockchain technology. This is consistent with prior research on incentivizing and rewarding consumers with cryptocurrency (Steinmetz et al., 2021). Thus, giving Explorers financial incentives in the form of cryptocurrencies could boost adoption.

With regards to promising blockchain applications, organizations should focus on self-sovereign identity and tokenization of assets. Their usefulness is currently held to be the highest. Business models building upon self-sovereign identity applications need to appeal to a customer group that is driven by optimism, trust in blockchain technology and which sees blockchain applications as bringing benefits to society. However, privacy concerns are relevant and need to

be alleviated. A similar notion applies for tokenization of assets applications, except that instead of privacy concerns, social influence is a driving factor. Therefore, managers in the context of tokenization applications need to be aware of and leverage the importance of network effects in growing their business. See, for instance, the current hype around NFT-collections. Explorers and Pioneers are most suitable for these applications. Anonymous transaction applications are perceived less useful. Yet, as the significance of trust and optimism is lower, managers should focus on appealing to Explorers for this application. Consumers perceive smart contract applications also to be less useful. This appears to be mainly driven by higher privacy concerns. It could be that consumers consider smart contracts as only containing highly confidential information, such as digital employment contracts or rental agreement contracts. Therefore, it is crucial for managers to ensure transparency and third-party verification of the functionality of smart contracts. Micropayment business models have been credited as one of the most likely upcoming blockchain developments (Schlecht et al., 2021). However, our results cast doubt on this assessment. We encourage managers to allow for a testing phase for micropayments in which consumers have time to get used to this new type of business model. From a consumer's perspective, fractional ownership applications score lowest on usefulness. However, managers can address similar customer groups as for tokenization applications.

7.3. Limitations and future research

Before drawing generalized conclusions from the results of this study, some considerations should be made. Generalizations of our findings might be limited to the German population – with respect to the field report, the British population. Consumers of other countries with different cultures are likely to have experienced a different socialization, which ultimately impacts their technology perception (Blut et al., 2022). Future studies should consider implementing cultural variables as moderators or conducting similar studies in other countries and regions to enhance cross-contextualization of our findings (Blut et al., 2022). Additionally, qualitative insights on contextual factors could be enhanced by conducting interviews. Furthermore, our survey-based research design has methodological limits. To measure actual behavior, future research should conduct experiments or field studies on user behavior. Note that clusters were designed based on TRI scores, which are technology independent. Cluster design is therefore free from blockchain-specific indicators. Our study calls for a more differentiable view at blockchain usage intention and blockchain applications. Future papers should examine the business model potential of blockchain applications that consumers find useful.

8. Conclusion

In this paper, we examine blockchain usage intention and application usefulness from the perspective of the consumer by conducting a quantitative study. We refine UTAUT-, TRI-, and blockchain specific theory and reveal which predictors are relevant in the context of blockchain adoption. Our research suggests several implications for practitioners, particularly with regards to fostering blockchain usage intention and assessing specific blockchain applications that look promising from the perspective of the consumer. However, in a highly dynamic market environment with surges in blockchain deal volumes and company valuations, forecasting the development of blockchain technology and its adoption is difficult. Consumer-centric research is required to examine the business model potential of blockchain applications. This enables businesses and consumers to gain a more profound understanding of the value potential of blockchain applications while ensuring that technology innovation and consumer perception are aligned.

References

Abdollahi, A., Sadeghvaziri, F., & Rejeb, A. (2022). Exploring the role of blockchain technology in value creation: a multiple case study approach.

Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. , *9*(2), 204–215.

Ali, O., Jaradat, A., Kulakli, A., & Abuhalimeh, A. (2021). A comparative study: Blockchain technology utilization benefits, challenges and functionalities. , *9*, 12730–12749.

Amberdata. (2022). *DeFi and the transformation of institutional finance.*

Aydiner, A. S. (2021). New approach to a disruptive business model with dynamic capability under the blockchain technology. In H. Dincer & S. Yüksel (Eds.), *Management strategies to survive in a competitive environment* (pp. 17–32). Springer International Publishing. (Series Title: Contributions to Management Science)

Babich, V., & Hilary, G. (2020). OM forum—distributed ledgers and operations: What operations management researchers should know about blockchain technology. , *22*(2), 223–240.

Blut, M., Chong, A. Y. L., Tsigna, Z., & Venkatesh, V. (2022). Meta-analysis of the unified theory of acceptance and use of technology (UTAUT): Challenging its validity and charting a research agenda in the red ocean. , *23*(1), 13–95.

Blut, M., & Wang, C. (2020). Technology readiness: a meta-analysis of conceptualizations of the construct and its impact on technology usage. , *48*(4), 649–669.

Bossler, L. F., & Kroenung, J. (2022). Exploring the current state of research on blockchain and cryptocurrency – analyzing enablers, inhibitors, and indeterminate factors. In *Wirtschaftsinformatik 2022 proceedings* (Vol. 4).

Boston Consulting Group. (2019). *Capturing the value of blockchain.*

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.).

Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. , *361*, 379.

Canton of Aargau. (2022). *Confirmation of residence.* https://ssi-poc.ag.ch.

Catalini, C., & Gans, J. S. (2016). Some simple economics of the blockchain. , *63*(7), 80–90.

CB Insights. (2021). *State of Blockchain.*

Chong, A. Y. L., Lim, E. T. K., Hua, X., Zheng, S., & Tan, C.-W. (2019). Business on chain: A comparative case study of five blockchain-inspired business models. , *20*(9), 1308–1337.

Cohen, B. H. (2013). *Explaining psychological statistics* (4th ed.). Hoboken, New Jersey, Wiley.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, N.J, L. Erlbaum Associates.

CoinGecko. (2022). https://www.coingecko.com/de. (Accessed: 16.04.2022)

Cong, L. W., & He, Z. (2019). Blockchain disruption and smart contracts. , *32*(5), 1754–1797.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. , *16*(3), 297–334.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. , 319–340.

Fishbein, M., Ajzen, I., & Belief, A. (1975). *Intention and behavior: An introduction to theory and research.* Addison-Wesley, Reading, MA.

Frizzo-Barker, J., Chow-White, P. A., Adams, P. R., Mentanko, J., Ha, D., & Green, S. (2020). Blockchain as a disruptive technology for business: A systematic review. , *51*, 102029.

Gefen, D. (2000). E-commerce: the role of familiarity and trust. , *28*(6), 725–737.

Hair, J. F. (Ed.). (2014). *Multivariate data analysis* (7th ed.). Harlow, Pearson.

Hawlitschek, F., Teubner, T., & Weinhardt, C. (2016). Trust in the sharing economy. , *70*(1), 26–44.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach* (2nd ed.). New York, Guilford Press.

Hendershott, T., Zhang, X. M., Zhao, J. L., & Zheng, Z. E. (2021). FinTech as a game changer: Overview of research frontiers. , *32*(1), 1–17.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. , *43*(1), 115–135.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. , *6*(1), 1–55.

Iansiti, M., & Lakhani, K. (2017). The truth about blockchain. , *95*, 118–127.

Ilk, N., Guangzhi, S., Shaokun, F., & Zhao, J. L. (2021-06). Stability of transaction fees in bitcoin: A supply and deemand perspective. , *45*(2), 563–592.

Jokisch, M. R., Schmidt, L. I., Doh, M., Marquard, M., & Wahl, H.-W. (2020). The role of internet self-efficacy, innovativeness and technology avoidance in breadth of internet use: Comparing older technology experts and non-experts. , *111*, 106408.

Kim, S. (2020). Fractional Ownership, Democratization and Bubble Formation - The Impact of Blockchain Enabled Asset Tokenization. In *Americas Conference on Information Systems 2020 Proceedings.*

Klöckner, M., Schmidt, C. G., & Wagner, S. M. (2022). When Blockchain Creates Shareholder Value: Empirical Evidence from International Firm Announcements. , *31*(1), 46–64.

Knauer, F., & Mann, A. (2020). What is in It for Me? Identifying Drivers of Blockchain Acceptance among German Consumers. , *3*(1), 1–16.

Koohikamali, M., Gerhart, N., & Mousavizadeh, M. (2015). Location disclosure on LB-SNAs: The role of incentives on sharing behavior. , *71*, 78–87.

Li, J. (2020). Blockchain Technology Adoption: Examining the Fundamental Drivers. In *Proceedings of the 2020 2nd International Conference on Management Science and Industrial Engineering* (pp. 253–260). Osaka, Japan, ACM.

Li, Y. (2014). The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns. , *57*, 343–354.

Liang, T.-P., Kohli, R., Huang, H.-C., & Li, Z.-L. (2021). What Drives the Adoption of the Blockchain Technology? A Fit-Viability Perspective. , *38*(2), 314–337.

Loh, X.-M., Lee, V.-H., Tan, G. W.-H., Ooi, K.-B., & Dwivedi, Y. K. (2020). Switching from cash to mobile payment: what's the hold-up? , *31*(1), 376–399.

Lu, Y., Zhao, L., & Wang, B. (2010). From virtual community members to C2C e-commerce buyers: Trust in virtual communities and its effect on consumers' purchase intention. , *9*(4), 346–360.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *The sage handbook of quantitative methodology for the social sciences* (pp. 175–198). Sage.

Marikyan, D., Papagiannidis, S., Rana, O., & Ranjan, R. (2021). Blockchain in a business model: Exploring benefits and risks. In *Responsible AI and analytics for an ethical and inclusive digitized society* (pp. 555–566). Cham, Springer International Publishing. (Series Title: Lecture Notes in Computer Science)

Marikyan, D., Papagiannidis, S., Rana, O. F., & Ranjan, R. (2022). Blockchain adoption: A study of cognitive factors underpinning decision making. , *131*, 107207.

McClelland, G. H., Irwin, J. R., Disatnik, D., & Sivan, L. (2017). Multicollinearity is a red herring in the search for moderator variables: A guide to interpreting moderated multiple regression models and a critique of iacobucci, schneider, popovich, and bakamitsos (2016). , *49*(1), 394–402.

Meyer, E., Welpe, I. M., & Sandner, P. (2021). Decentralized Finance—A systematic literature review and research directions.

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system.* https://bitcoin.org/bitcoin.pdf.

Nowiński, W., & Kozma, M. (2017). How Can Blockchain Technology Disrupt the Existing Business Models? , *5*(3), 173–188.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, McGraw-Hill.

Olsen, T. L., & Tomlin, B. (2020). Industry 4.0: Opportunities and Challenges for Operations Management. , *22*(1), 113–122.

Parasuraman, A. (2000). Technology Readiness Index (Tri): A Multiple-Item Scale to Measure Readiness to Embrace New Technologies. , *2*(4), 307–320.

Parasuraman, A., & Colby, C. L. (2015). An updated and streamlined technology readiness index: TRI 2.0. , *18*(1), 59–74.

Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: Integrating trust and risk with the technology acceptance model. , *7*(3), 101–134.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. , *88*(5), 879–903.

Quiroz-Gutierrez, M. (2022). Bored apes and CryptoPunks help jolt NFT market to over 21,000% growth and $17.6 billion in sales last year.

Raddatz, N., Coyne, J., Menard, P., & Crossler, R. E. (2021). Becoming a blockchain user: understanding consumers' benefits realisation to use blockchain-based applications. , 1–28.

Rogers, E. M. (1962). *Diffusion of innovations*. Free Press.

Rossi, M., Mueller-Bloch, C., Thatcher, J. B., & Beck, R. (2019). Blockchain research in information systems: Current trends and an inclusive future research agenda. , *20*(9), 1388–1403.

Rönkkö, M., & Cho, E. (2022). An updated guideline for assessing discriminant validity. , *25*(1), 6–14.

Sarker, S., Henningsson, S., Jensen, T., & Hedman, J. (2021). Use of blockchain as a resource for combating corruption in global shipping: An interpretive case study. , *38*(2), 338–373.

Schlecht, L., Schneider, S., & Buchwald, A. (2021). The prospective value creation potential of blockchain in business models: A delphi study. , *166*, 120601.

Steinmetz, F., von Meduna, M., Ante, L., & Fiedler, I. (2021). Ownership, uses and perceptions of cryptocurrency: Results from a population survey. , *173*, 121073.

Toufaily, E., Zalan, T., & Dhaou, S. B. (2021). A framework of blockchain technology adoption: An investigation of challenges and expected value. , *58*(3), 103444.

Venkatesh, Morris, Davis, & Davis. (2003). User acceptance of information technology: Toward a unified view. , *27*(3), 425.

VISA. (2021). *The Crypto Phenomenon: Consumer Attitudes & Usage.*

Warshaw, P. R., & Davis, F. D. (1985). Disentangling behavioral intention and behavioral expectation. , *21*(3), 213–228.

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. , *26*(2), 13–23.

Weking, J., Mandalenakis, M., Hein, A., Hermes, S., Böhm, M., & Krcmar, H. (2020). The impact of blockchain technology on business models – a taxonomy and archetypal patterns. , *30*(2), 285–305.

Whitaker, A., & Kräussl, R. (2020). Fractional Equity, Blockchain, and the Future of Creative Work. , *66*(10), 4594–4611.

Yin, H. H. S., Langenheldt, K., Harlev, M., Mukkamala, R. R., & Vatrapu, R. (2019). Regulating cryptocurrencies: A supervised machine learning approach to de-anonymizing the bitcoin blockchain. , *36*(1), 37–73.

Zhang, W., Wei, C.-P., Jiang, Q., Peng, C.-H., & Zhao, J. L. (2021). Beyond the block: A novel blockchain-based technical model for long-term care insurance. , *38*(2), 374–400.

Zheng, Y., & Boh, W. F. (2021). Value drivers of blockchain technology: A case study of blockchain-enabled online community. , *58*, 101563.

Ziolkowski, R., Miscione, G., & Schwabe, G. (2020). Decision Problems in Blockchain Governance: Old Wine in New Bottles or Walking in Someone Else's Shoes? , *37*(2), 316–348.