Doctoral Thesis

Essays in Behavioral Analytics
Using Data Science and Machine Learning

Author Hannes Gerstel

Submitted to Justus-Liebig-University Giessen
School of Business and Economics

Submission date July 19, 2023

Supervisors

Prof. Dr. Arnt Wöhrmann
Chair of Managerial Accounting

Prof. Dr. Corinna Ewelt-Knauer
Chair of Financial Accounting

Table of Contents

Chapter	Project Title
I	Does the Devil's Advocate Approach Mitigate Escalation of Commitment?
II	Celebrating Failure – The Effects of Failure Awards on Risk-Taking and Escalation of Commitment
III	What Motivates Independent Directors? The Influence of Director Incentives on Director Decisions
IV	Do Opposites Attract? The Effect of Cultural Distance on Mergers and Acquisitions: Evidence from Glassdoor Reviews
V	Machine Learning vs. Management Forecasts: Können Machine-Learning-Modelle die Genauigkeit von Umsatzprognosen verbessern?

Does the Devil's Advocate Approach Mitigate Escalation of Commitment?

Authors Hannes Gerstel, Niklas Kreilkamp, Maximilian Schmidt, Arnt Wöhrmann

Abstract Using an experimental setting, we examine under which circumstances the devil's advocate method reduces decision-makers' tendencies to continue poorly performing projects, also known as escalation of commitment (EoC). We posit that decision-makers' willingness to process the devil's advocate's critique properly depends on a firm's organizational error management climate (EMC). Due to contradicting prior research findings, we first analyze whether open (learning from errors) vs. blame (preventing any error from occurring) EMCs impact escalation tendencies. Process evidence reveals a two-sided effect of the open EMC, resulting in similar escalation levels as in blame EMCs. Moreover, we examine the effectiveness of the devil's advocate in the open vs. blame EMC. Using prominent EoC drivers, we show that implementing a devil's advocate reduces escalating behavior in both EMCs. Our findings have practical relevance for firms implementing a devil's advocate as an effective management control to prevent project escalation.

Publication Status Revise & Resubmit

Journal of Business Economics (VHB-Jourqual 3: B)

Workshop/ Conference Presentations

2019 Doctorial Colloquium on Experimental Management Accounting Research (Ruhr University Bochum)

2019 Summer School on Experimental Research in Management Accounting (University of Bayreuth)

Acknowledgements We thank Alexander Brüggen, Marcel van Rinsum, Ivo Schedlinsky, Friedrich Sommer and Michael Williamson for their helpful comments.

Author Contribution	Gerstel	Kreilkamp	Schmidt	Wöhrmann
Relative Share	*75%*	*5%*	*5%*	*15%*
Conceptualization	✓	✓	✓	✓
Theory Development	✓			
Methodology	✓			
Data Acquisition	✓			
Data Analyses	✓			
Writing	✓			✓

1 Introduction

The success of investment projects determines corporate performance. Yet, managing these projects underlies decision-making processes prone to biases such as escalation of commitment (EoC) (Roetzel et al. 2020; Booth and Schulz 2004). EoC refers to situations in which decision-makers stick to a failing course of action, despite receiving negative feedback on their initial decision (Staw 1976), i.e., they continue investing in a poorly performing project. 70% of respondents in a survey study by Kreilkamp et al. (2021) state that EoC at least frequently occurs within their organization. To counter resource wastage, firms can apply several management controls and techniques that facilitate de-escalation (e.g., Fehrenbacher et al. 2020; Brüggen and Luft 2016).

In this vein, a common recommendation is the implementation of independent third-party reviews of running projects (e.g., Loh et al. 2019; Kadous and Sedor 2004). Herbert and Estes (1977) suggest the devil's advocate technique to challenge assumptions of corporate decision-making. Today's practitioner's view defines the devil's advocate as "a formal role assigned to an individual before a final decision is made [...] to challenge the current view of the decision-maker" (Hoffman et al. 2018, p. 8).

For example, the German energy company RWE started to formally appoint a devil's advocate to question the underlying assumptions of important decisions, after spending several billion euros on failing projects (Günther et al. 2017). Moreover, besides Royal Dutch Shell and IBM (Ivancevich et al. 2011), JPL NASA also formally appoints devil's advocates to make decision-makers more aware of potential defects in their decision-making processes (Kaplan and Mikes 2012). Further, Google applies a similar technique where project members are appointed to highlight negative information indicating project failure (Google Rework 2023).

There are different suggestions concerning the specific design of the devil's advocate role (Schwenk 1988; Cosier 1981). However, it is univocal that the devil's advocate must highlight negative information contradicting the proposed decision. In case of mitigating escalating behavior, this implies highlighting negative information (that questions further investments in a poorly performing project). Nevertheless, the devil's advocate technique sometimes involves additional processes, e.g., collecting new information or giving a specific recommended course of action (Schwenk 1984). To increase generalizability, we build on the unambiguous feature of highlighting negative information and refrain from other inconclusive features of the devil's advocate.

Research examining whether the devil's advocate mitigates EoC is scant and inconclusive. For example, Greitemeyer et al. (2009) find that a devil's advocate does not generally reduce EoC in a group setting. De-escalation is only observed in the case of multiple continuation decisions and if there is already dissent within the management team concerning the investment policy. Schwenk (1988) examines a setting where an external advisor (erroneously) suggests funding a failing project. He finds that while the devil's advocate marginally reduces the investment triggered by the expert opinion, it does not offset it.

These findings raise the question of why there is no compelling evidence for the effectiveness of the devil's advocate as a de-escalation tool. The survey results of Kreilkamp et al. (2021) provide a possible explanation by showing that the effectiveness of debiasing techniques depends on a firm's cultural controls. They find that without a certain level of psychological safety (Edmondson 1999), debiasing techniques are rendered ineffective. Psychological safety –the feeling of being safe for interpersonal risk-taking– is a major component of a firm's cultural controls, which is influenced by a firm's organizational error management climate (EMC). EMCs can generally be classified on a continuum from open to blame (Gold et al. 2014; Edmondson 2002). Organizations with a blame EMC emphasize sanctions, punishment and blaming to prevent all errors from occurring (Frese and Keith 2015). On the contrary, in an open EMC, errors are dealt with in an active and open manner to reduce negative error consequences and increase positive effects. In an EoC scenario, terminating a project implies admitting it failed (Harrell and Harrison 1994).

Due to mixed evidence from prior research on the effect of EMC on EoC (Fischer et al. 2018; Mahlendorf 2015; Barton et al. 1989), we posit a research question for the main effect of EMC on escalation (*RQ1*). On the one hand, an open EMC might encourage exploration and risk-taking (Fischer et al. 2018; Edmondson 1999). In the case of a poorly performing project, decision-makers might bet on the small chance of turning the project around, since they do not expect any repercussions if the endeavor fails. Thereby, an open EMC could increase escalating behavior.

On the other hand, an open EMC can induce the feeling of being safe for interpersonal risk-taking (Cannon and Edmondson 2005, 2001). Consequently, decision-makers are more likely to admit project failure and thus are more likely to terminate their project. Supporting the de-escalating effect of an open EMC, Mahlendorf (2015) shows that higher organizational "allowance for failure" reduces project managers' perceived threat of project failure. He links the lower perceived threat of project failure to reduced escalating behavior. Supporting the escalating effect of an open EMC, Barton et al. (1989) find that project escalation increases when decision-makers are (wrongfully) assured that their initial decision was based on good judgement.

Since EoC is based on the premise of a poorly performing investment, decision-makers' willingness to accept project failure and properly process the devil's advocate's critique will likely depend on a firm's EMC. Thus, we examine the effectiveness of the devil's advocate technique in an open (*H1*) vs. blame (*RQ2*) EMC. We argue that the devil's advocate reduces decision-makers' tendencies to focus on goal completion (Conlon and Garland 1993) and their tendency to overweigh positive information (Caldwell and O'Reilly 1982). We expect that decision-makers in an open EMC are willing to admit project failure after processing the highlighted negative information, leading to lower levels of EoC (*H1*). On the contrary, decision-makers in a blame EMC refrain from admitting mistakes to avoid public embarrassment (Gronewold et al. 2013). Highlighted negative information indicating project failure could evoke the perception to the decision-maker that it was a mistake to start the project in the first place. We argue that the effectiveness of the devil's advocate in a blame EMC depends on the project owners' perceptions on whether they classify starting an initially promising project – which is now poorly performing– as a mistake. If decision-makers correctly classify project initialization as the right decision and not a mistake, they are more willing to admit failure of the project. However, due to an environment of punishment and blame, decision-makers might refrain from openly admitting project failure in a blame EMC after processing the highlighted negative information, rendering the devil's advocate ineffective (*RQ2*).

To test our predictions, we employ a 2×2 between-subjects experiment and measure participants' recommendations to continue a poorly performing project as our proxy for EoC. Similar to Gold et al. (2014), we manipulate EMC at *open* (learning from mistakes) vs. *blame* (preventing any error from occurring). The devil's advocate, who provides highlighted negative information, is also manipulated at two levels (*present* versus *absent*) while keeping the information constant in all conditions. In our experimental task, participants act as R&D managers for a fictitious company. First, they must select one of two R&D projects to start. Afterward, they receive negative feedback indicating an unforeseen decline in expected returns of their initiated project. Participants then have to give a recommendation to the management board on whether to continue the project or terminate it and invest the remaining funds in an economically superior alternative.

Our experimental data reveals several important results. First, we do not find a main effect of the EMC on EoC (*RQ1*). Additional analyses reveal a two-sided impact of the open EMC on EoC (being able to admit project failure vs. increased risk-taking and not being concerned about repercussions), explaining the insignificant EMC main effect. Second, we find that the presence of the devil's advocate decreases participants' escalating behavior in the open (*H1*) and blame (*RQ2*) EMC. Process evidence reveals that assigning a person to criticize the poorly performing project sends a clear signal to the decision-maker that the goal of project completion should be reassessed, independent of the EMC. Furthermore, decision-makers' tendency to overweigh positive information is successfully attenuated by the devil's advocate in both EMCs.

Our study makes several contributions to theory and practice. First, we show that the devil's advocate attenuates EoC through the simple and time-efficient process of highlighting negative information. In contrast to prior research, which implements a third-party giving a recommendation (Behrens and Ernst 2014; Schwenk 1988), our de-escalating process does not require a recommendation or interpretation of information. Both leave room for false recommendations and misinterpretation of information (from external parties), which is not an issue when implementing a devil's advocate with the assignment to highlight existing negative information. Our findings are particularly informative for firms such as Google, Facebook, RWE, and others that increasingly use debiasing techniques (Baer et al. 2017; Facebook 2022; Google Rework 2023). Our results show that providing highlighted negative information effectively reduces EoC in open and blame EMCs by lowering decision-makers' susceptibility to the goal-substitution effect (Conlon and Garland 1993) and their overweighing of positive information (Caldwell and O'Reilly 1982).

It should be noted that highlighting negative information does not necessarily require a formal devil's advocate. Instead, the task could be (informally) assigned to other parties, e.g., management accountants. However, it is crucial that the assessor has a general understanding of the project's circumstances (Kadous and Sedor 2004). Furthermore, the assessor should not have any stake in the project (e.g., personal connections or financial interests), since Fehrenbacher et al. (2020) find that independent assessors tend to support poorly performing projects if they exhibit an affective reaction toward the project manager.

Second, we contribute to the broad (accounting) literature examining EoC. Prior literature examines and recommends several de-escalation tools. For example, shifting decision-makers in escalation situations (Brüggen and Luft 2016; Sleesman et al. 2012), using the real options approach (Denison 2009), and setting objective target goals before spending further resources on projects (Zimmerman 2019; Cheng et al. 2003) should lead to de-escalation. However, such de-escalation tools also come with potential caveats. Brüggen and Luft (2016) find that rotating project managers leads to highly optimistic budget forecasts of their subordinates, resulting in cost overruns. Cheng et al. (2003) find that decision-makers still escalate after firms set a minimum rate of return targets (hurdle rate). Even though self-set hurdle rates effectively reduce EoC, the project owner still has leeway to adjust the hurdle rate to her own interest. Similar to the finding of Denison (2009) that a real options approach decreases EoC, the real options approach is also prone to subjectivity when generating (future-based) key inputs (Trigeorgis 1996).

On the contrary, our process evidence shows that the de-escalating strategy of providing highlighted negative information directly debiases the project owner's decision-making process. This has the advantage of directly keeping the expertise of the project owner in the decision process. We add to the findings of Loh et al. (2019), who show that external consultants are less likely to recommend continuation of a poorly performing project than internal consultants. By taking the project owner's perspective, we show that highlighted negative information of external consultants significantly lowers EoC tendencies. Lastly, psychology literature often implements the devil's advocate exclusively in group settings (to combat groupthink biases) (Greitemeyer et al. 2009; Schwenk and Cosier 1993). Yet, many decisions are made by an individual (Plous 1993). For example, in publicly listed firms, the CEO is usually the one who ultimately decides (Feld 2011). Thus, we also contribute to the devil's advocate literature by showing its effectiveness in a scenario which is more representative of an individual decision-making scenario.

Third, our findings add to research investigating the direct effect of EMCs on escalation. Prior research finds mixed evidence concerning the impact of the EMC on EoC (Fischer et al. 2018; Mahlendorf 2015; Barton et al. 1989). Survey results of Mahlendorf (2015) indicate that an open EMC reduces escalation. Fischer et al. (2018) counter that an open EMC could encourage experimentation, risk-taking, and ultimately escalation. Our study addresses the potential mono-method bias of prior research. We add experimental and process evidence, offering an explanation to the survey results of Mahlendorf (2015) and the assumptions of Fischer et al. (2018).

The remainder of this paper proceeds as follows. Section 2 presents the theoretical background and hypotheses. Section 3 outlines our research design. Section 4 presents and discusses our results. Section 5 concludes.

2 Theoretical Background and Hypotheses Development

2.1 Organizational Error Management Climate and Escalation of Commitment (*RQ1*)

Before developing our first hypothesis (*H1*) and second research question (*RQ2*) on how the organizational error management climate (EMC) determines the effectiveness of the devil's advocate technique, we derive our first research question (*RQ1*) on the main effect of EMC on escalation of commitment (EoC).

EoC was first examined by Staw (1976), who finds that personal responsibility of project managers leads to the irrational continuation of poorly performing projects. Staw and Ross

(1987) developed an EoC framework and identified four sets of determinants (project, psychological, social, and structural determinants). Since the devil's advocate directly interferes with the project owner's decision-making process, we focus on psychological determinants to derive the impact of the devil's advocate on EoC (*H1* and *RQ2*). For the impact of the EMC on EoC (*RQ1*), we use social determinants since the EMC mainly influences the decision-maker's surroundings.

Social determinants are reflected by corporate culture (Sleesman et al. 2012; Staw and Ross 1987). One key aspect of corporate culture is reflected by a firm's EMC (Klein et al. 1994), encompassing firms' practices and procedures for dealing with errors (van Dyck et al. 2005). A firm's EMC can be classified on a continuum from open (learning from errors) to blame (preventing any error from occurring) (Gronewold et al. 2013; van Dyck et al. 2005). Errors (deviations from an expected or intended outcome) often lead to negative consequences and are generally seen as a signal of poor performance. Organizations with a blame EMC build upon this perspective and try to avoid errors by emphasizing sanctions, punishment, and blaming. Hence, admitting an error in blame EMC may cause financial and personal distress, such as reputation loss (Zhao and Olivera 2006).

In this vein, Gronewold et al. (2013) find that auditors are less willing to admit errors in a blame EMC, as mistakes are seen as a signal of incompetence. In the case of EoC, terminating a project implies admitting its failure (Harrell and Harrison 1994). Consequently, decision-makers are afraid of admitting and explaining their mistake and being perceived as incompetent, leading to escalating behavior. In line with this argument, survey results by Mahlendorf (2015) indicate that low organizational "allowance for failure" increases the perceived threat of project failure. In contrast, a high "allowance for failure" refers to the notion that failure is an acceptable outcome for projects involving risk factors. Thus, a lower allowance for failure increases EoC, as project managers fear the negative consequences of project termination (e.g., reputational damage).

On the contrary, an open EMC implies that a firm deals with errors in an active and open manner (Frese and Keith 2015; van Dyck et al. 2005). Thereby, the negative feeling of being afraid of personal consequences after admitting an error are reduced or even eliminated. An open EMC also comprises positive social reactions to errors, e.g., sharing error knowledge and promoting learning from errors.

These arguments indicate higher levels of EoC in blame compared to open EMCs. However, the overall effect of a firm's EMC on escalating behavior remains unclear. EMC has commonalities with the construct of psychological safety, which is defined as individuals' beliefs to be safe for interpersonal risk-taking within their workplace (Cannon and Edmondson 2005; Edmondson 2002, 1999). Operating in an environment with characteristics of an open EMC can induce the feeling of psychological safety. However, the overall effect of psychological safety on EoC is less univocal (Edmondson and Lei 2014; Baer and Frese 2003).

On the one hand, high psychological safety enables individuals to make decisions (e.g., terminating a poorly performing project) without fearing negative consequences, such as a reputation loss or a demotion (Kahn 1990). Moreover, accepting (negative) feedback and changing one's opinion is not perceived as a weakness (Edmondson 1999). Thus, psychological safety increases the willingness to admit project failure (O'Neill 2009). Consequently, a psychologically safe environment could lead to less escalation. Prior research finds that framing negative project outcomes as less threatening (Simonson and Staw 1992; Heng et al. 2003), facilitating open discussions, and implementing collective responsibility (Shepherd et al. 2019; Edmondson and Lei 2014; Pan and Pan 2011) –which can be seen as components of an open EMC– have a mitigating effect on EoC.

In contrast to the potential de-escalating effect due to feeling safe to admit (project) failure, an open EMC may also facilitate escalation. Edmondson (1999) argues that an open EMC fosters the feeling of being safe to experiment, which encourages risk-taking. Since decision-makers do not anticipate blame, they are protected from negative consequences (Newman et al. 2017). Fischer et al. (2018) also postulate that an open EMC encourages exploration and risk-taking, which would facilitate EoC. In the case of a poorly performing project, decision-makers might bet on the small chance of turning the project around, since they do not expect any repercussions if the endeavor fails. Therefore, an open EMC could also facilitate escalating behavior. Based on the contradicting effects of an open EMC on EoC –feeling safe to admit failure vs. taking risks due to not expecting repercussions– we derive our first research question:

***RQ1*: Do decision-makers exhibit lower levels of escalation of commitment in an open compared to a blame error management climate?**

2.2 The Devil's Advocate in an open versus blame Error Management Climate (*H1* & *RQ2*)

Next, we derive the effects of the effectiveness of the devil's advocate technique in an open (*H1*) versus a blame EMC (*RQ2*). In the case of a failing project, the de-escalating effect of the devil's advocate materializes via a two-step process: First, the provided highlighted negative information by the devil's advocate must change the decision-maker's assessment of project success (*change of assessment*). Second, the decision-maker must be willing to admit project failure and stop funding the project (*translation into action*).

Referring to the first step, individuals usually seek confirming (positive) information while discounting negative information (Hogarth 1987; Caldwell and O'Reilly 1982). When the devil's advocate highlights negative information, the risk of failure becomes salient. In this vein, Denison (2009) finds that using a real options model for an investment appraisal –instead of a net present value approach– decreases EoC. This occurs because the mental accessibility of the possibility of abandoning the poorly performing project increases. Similarly, we argue that highlighted negative information makes potential project failure more salient. Thereby, the concept of abandoning the project also becomes more salient. The more accessible a cognitive concept is, the more likely it is to influence individuals (Fazio et al. 1986). In line with this expectation, Brockner et al. (1981) find that supplying individuals with highlighted cost information decreases their tendency to become entrapped during a waiting task. In the experimental task of Brockner et al. (1981), participants believed they could win a prize at a randomly chosen time while costs incurred for waiting. By displaying these costs, individuals became more attentive to the costs and became less willing to escalate. Further, Behrens and Ernst (2014) observe that EoC is reduced if project managers receive a visualization of already-known information.

Moreover, as part of psychological EoC determinants (Sleesman et al. 2012), the goal-substitution effect is a prominent EoC driver. It describes the phenomenon that the closer a project gets to completion, the more emphasis decision-makers put on completing the project rather than on achieving their original goals (e.g., economic success) (Conlon and Garland 1993). Making decision-makers aware of the costs and risks associated with project continuation –by applying the devil's advocate technique– decreases their perceived proximity to (successful) project completion. Thereby, termination becomes more likely.

Concerning the first process step, we conclude that the devil's advocate technique affects the assessment of project success in both EMCs by reducing the psychological determinants to overweigh positive information and putting too much emphasis on irrational project completion. Therefore, a decision-maker is more likely to form realistic profit expectations when highlighted negative information is provided. However, it is not clear whether the second step –of being willing to admit project failure and stop funding the project– can effectively materialize in both EMCs.

Open EMCs exhibit high levels of psychological safety (Cannon and Edmondson 2005; Edmondson 2002). In this environment, employees do not fear engaging in constructive conflict and confrontation (Edmondson 1999). As a result, decision-makers do not feel threatened by criticism of their decisions and appreciate the information provided by the devil's advocate, even though it indicates project failure. Nevertheless, an open EMC can also encourage exploration and risk-taking (Fischer et al. 2018; Edmondson 1999). Consequently, decision-makers might emphasize the opportunity to turn the poorly performing project around. Yet, their willingness to take this risk depends on whether the situation is perceived as a threat or an opportunity (Highhouse and Yüce 1996). By implementing the devil's advocate technique, the firm sends a clear signal to the decision-maker that continuing poorly performing projects is classified as a threat and that wasting resources on such projects is not desirable. For open EMCs, we argue that decision-makers are less inclined to take the risk of project continuation due to the clear threat signal sent by the firm.

Taken together, we posit that the first (changing the assessment of project success) and second requirement (being willing to admit project failure) are met in the open EMC, resulting in a de-escalating effect of the devil's advocate. This leads to the following hypothesis:

***H1*: The devil's advocate reduces escalation of commitment in an open error management climate.**

Organizations with a blame EMC try to avoid errors by emphasizing punishment and blaming, e.g., by decreasing a decision-maker's responsibility for future projects (van Dyck et al. 2005). Highlighted negative information indicating project failure could evoke the perception to the decision-maker that it was a mistake to start the project in the first place. Mistakes are defined as deviations from an expected or intended outcome (Reason 2000). Objectively seen, the initial decision to start an economically promising project does not classify as a mistake, while continuing the poorly performing project objectively does (Harrison and Harrell 1993). Irrespective of justified prior decision-making, blame EMCs tend to establish a general perception of bad outcomes being a mistake (Gold et al. 2014), e.g., admitting failure of an initially promising project.

We argue that the effectiveness of the devil's advocate in a blame EMC depends on the project owners' perceptions of whether they classify starting an initially promising project – which is now poorly performing– as a mistake. On the one hand, if decision-makers correctly classify project initialization as the right decision and not a mistake, they are more willing to admit project failure. Consequently, the devil's advocate technique would make termination more likely. On the other hand, due to an environment of punishment and blame, decision-makers might refrain from admitting failure in a blame EMC. By continuing the project instead of admitting the poor performance, the de-escalating effect of the devil's advocate would not materialize. Supporting this line of thought, Gronewold et al. (2013) find that auditors are less willing to admit errors in a blame EMC, where mistakes signal a lack of competence.

Taken together, we argue that the first step (change of assessment) of the devil's advocate technique also materializes in blame EMCs. However, it is less clear whether decision-makers are willing to admit project failure after processing the highlighted negative information. Thus, we phrase the impact of the devil's advocate as a research question:

***RQ2*: Does the devil's advocate reduce escalation of commitment in a blame error management climate?**

3 Research Method

3.1 Experimental Design and Procedures

To test our hypotheses and answer our research question, we employ a 2 × 2 between-subjects experimental design using the experimental software oTree (Chen et al. 2016). We manipulate *the devil's advocate* at two levels (*devil's advocate absent* vs. *present)* and *EMC* at two levels (*open* vs. *blame)*. Our dependent variable is the participants' decision to continue a poorly performing R&D project. The experiment consists of two parts. In the first part, we collected data on participants' demographics and risk preferences. The second part is our main task which entails the manipulations and the escalation task. The second part was conducted two weeks after the first part to reduce the time load of the second part. Figure 1 depicts an overview of the experimental procedure. Participants conducted both parts from home. We implemented several mechanisms to mitigate potential concerns related to non-laboratory experiments (Charness et al. 2007). First, we implemented an incentive scheme in which participants' decisions directly influenced their payout. Second, we prohibited mobile device use and prompted participants to use their web browsers' full-screen mode to minimize distractions. Third, we measured participants' time spent on each screen to check their attentiveness. Fourth, participants had to open all reading sections of a page to be able to advance to the next page.[1]

[Insert Figure 1 about here]

We measured risk preferences in the first part using a lottery choice task (Sprinkle et al. 2008). During this task, participants have to indicate for 15 scenarios whether they prefer a safe payment (0.75€) or wish to participate in a lottery. Depending on the scenario, the lottery either pays 1.50€ with a probability of p, or 0€ with a probability of (1-p). The probability p decreases from 85% (scenario 1) to 15% (scenario 15) in 5% increments. Since the expected value of scenario 8 is equal to the safe payment, participants switching to the safe payment before (at; after) scenario 8 are classified as risk-averse (risk-neutral; risk-seeking). After the lottery, participants answered several demographic and personality items.

[1] Participants took 41.32 minutes to complete the experiment (median). Excluding outliers (+/- 2 standard deviations from the mean time spent on the main task) does not change our results.

The second part represents the main experiment, in which participants had to act as R&D managers. Participants read the instructions, learned their compensation, and the EMC was manipulated. After correctly answering several comprehension questions during a quiz, participants proceed to the main task. The experiment concluded with a post-experimental questionnaire. Afterward, participants were informed about their compensation outcome.

3.2 Main Task

The main task is similar to Seybert (2010) and Brink et al. (2020). Participants assumed the role of R&D managers at the fictitious company "Kitchen World". They had to make two decisions: First, they had to fund one of two projects. Second, they had to decide whether to provide further funding to the selected project (our proxy for EoC) after receiving negative project feedback.

For the first decision, participants had to choose one out of two project ideas for funding: a "Smart Coffeemaker" or a "Smart Teamaker".[2] Participants were informed about both projects' required investment and expected discounted cash inflows. Additionally, a brief product description and market information was provided. The required investment, the cash flows, and, thus, the NPV (4.5m lira) were identical and positive for both projects. [3] The qualitative information differed slightly between the two projects.

To induce an EoC setting during the main task, decision-makers have to receive negative feedback on their initial decision (Wong et al. 2006). Seybert (2010) and Denison (2009) provide negative information to their participants after the initial decision, which leads to a decline in expected cash flows. Similarly, our participants receive negative project feedback

[2] Since we want to observe escalating behavior, we intentionally let participants make the initial decision instead of letting the board chose the project or provide an inherited project (Brink et al. 2020). By letting participants decide, perceived involvement with the chosen project is induced, which fosters escalating behavior (Seybert 2010; Staw 1976).

[3] We used the fictitious experimental currency lira during the experiment.

after making the first investment decision. Participants learn that a competitor develops a similar product and that the planned cooperation with a distribution partner is at stake. If the project is continued, an additional investment of 3.5m lira is required. Based on updated forecasts, the project yields either 9.5m expected lira cash inflows in the best-case (33% probability) or 4.5m expected lira in the worst-case scenario (67% probability). The NPV of project continuation is $[0.33 \times 9.5\text{m lira} + 0.67 \times 4.5\text{m lira}] - 3.5\text{m lira} = 2.65\text{m lira}$, while the predicted NPV was 4.5m lira when the initial project was started. Participants have to assess whether they provide additional funding or terminate the project. In the latter case, the remaining 3.5m lira are invested in an alternative, yielding an NPV of 3.5m lira (= 7m lira cash inflows – 3.5m lira required initial investment). Hence, termination is the economically preferred option.

Participants in the devil's advocate condition received a brief explanation of the process of highlighting negative information, which was recently implemented by the executive board (Appendix 1). Afterward, they received an objective report covering the potential risks of project continuation. Finally, all participants were prompted to give their final recommendation regarding project continuation. Participants respond on a 101-point scale ranging from 0 (termination) to 100 (continuation) on how strongly they recommend project continuation to the management board (Appendix 2).[4] Participants learn that the value corresponds to the board's propensity to follow their recommendation. For example, a value of 80 implies that the board

[4] Contrary to the initial decision, which is made by the participant, we frame the continuation decision as a recommendation to the board for two reasons. First, to induce a direct cultural influence on the continuation decision, we introduce the management into the decision process using upper echelons theory (Hambrick 2007), which states that the organizational outcomes are culturally influenced by its top executives. Simultaneously, participants can assume that they actually make the subsequent decision by reading that "you can assume that the management board will follow your recommendation". Second, to ensure that participants take the decision seriously while also representing a realistic scenario, we directly link participants' compensation to their recommendation value by referring to the board's propensity to follow the recommendation.

will continue the project with a probability of 80%. We use participants' responses to this question to measure EoC, with higher values indicating a greater tendency to escalate.[5] A random mechanism using the probabilities from participants' recommendation values determines whether the project is continued. Next, using the given project outcome probabilities, another random mechanism determines whether the best- or worst-case materializes in case of project continuation. Thus, participants' recommendation values directly affected their compensation. Participants earned a performance-contingent compensation on the main task, which ranged between 4.50 and 9.50€.[6] The total average compensation was 11.68€. Additionally, they earned a show-up fee (4.25€) and compensation for the lottery task (0.75€ on average).

3.3 Devil's Advocate Manipulation

We manipulate our first independent variable at two levels (*devil's advocate* absent vs. present). In the devil's advocate present treatment, the devil's advocate was introduced right before the continuation decision. Participants were informed that Kitchen World's board had hired an external consulting firm to review the project.[7] The consultancy's sole assignment was to highlight threats that speak against project continuation, as they might lead to project failure (Appendix 1 Panel A).[8] Participants were informed that the consultancy was explicitly *not* asked

---

[5] To ensure that participants take the updated financial information seriously after receiving the negative feedback, we perform two additional analyses (untabulated). First, we remove the fastest 20^{th} percentile (25 participants) spending a maximum of only 30 seconds on the recommendation page. Second, we exclude participants indicating a value below 4 (39 participants) on the item "I mainly made my decision based on the given quantitative information" (1 = fully disagree; 7 = completely agree). For both analyses the main results do not change, indicating that participants took the provided information seriously.
[6] The variable renumeration in lira is calculated as follows:
$Variable\ renumeration\ in\ lira = Project\ account\ balance * 0.01$; with
$Project\ account\ balance = 7.5m\ lira + (realized\ cash\ inflows - investments)$
10,000 lira were converted to 1€.
[7] Since Loh et al. (2019) show that external consultants are more efficient in reducing escalating tendencies than in-house consultants, the highlighted negative information is provided by an external party.
[8] To ensure that the firm's EMC cannot be inferred based on the assignment of the devil's advocate, the description of the devil's advocate is stated as neutral as possible (Appendix 2, Panel A). We discuss a potential causality between the firm's EMC and its propensity to implement a devil's advocate in the conclusion.

for a recommendation or listing the project's opportunities. After participants read the consultancy's assignment, they received a brief report listing two threats that might lead to project failure (Appendix 1 Panel B).

To hold information constant across conditions and to increase internal validity, the consultancy only repeated threats that the participants were aware of in all treatments. No new information was provided. In the devil's advocate absent treatment, participants did not receive the consultancy's assignment or the report. However, participants in the absent treatment could also review the project information a second time before giving their continuation recommendation.

3.4 Error Management Climate Manipulation

We follow Gold et al. (2014) and manipulate *EMC* at two levels (open versus blame). According to prior research, a firm's EMC depends on several factors. For example, it is affected by whether errors are dealt with in an active and open manner, and the potential consequences of committing an error (Gold et al. 2014; Gronewold and Donle 2011; van Dyck et al. 2005). Appendix 3 shows the manipulation, and Table 1 Panel A highlights the differences between the EMC treatments. Participants had to correctly answer two quiz items to verify their understanding of the firm's EMC (Table 1 Panel B).[9]

[Insert Table 1 about here]

Participants were told that error management is a crucial aspect of *Kitchen World's* daily business. Based on the upper echelon's theory –stating that the organizational outcomes are culturally influenced by its top executives (Hambrick 2007)– and the fact that supervisors are often the carriers of culture (Zohar and Luria 2004), participants learn how the management board defines the firm's EMC. In the *blame* [open] EMC treatment, participants read: "The

[9] Removing participants from our analyses, who initially didn't correctly answer at least one of the two EMC quiz items, does not change our results (untabulated).

board is convinced that wrong decisions *are* [not] always avoidable and therefore attaches great importance to the fact that employees *do their utmost to avoid possible errors* [accept and learn from wrong decisions]." Further, participants read an e-mail from the management board to the marketing department that illustrates Kitchen World's EMC. In this e-mail, the board comments on a poorly performing marketing project. For example, in the *blame* [open] EMC, the marketing managers learn that the failure "*will* [not] have a negative impact on your annual performance evaluation."[10]

3.5 Participants

128 business students from a large Western European university participated in the experiment and were randomly assigned to one of the four treatments.[11] Participants were at least in their 2nd year of business studies. Thus, all participants were familiar with calculating expected values, the concept of net present values and sunk cost, and had a sufficient understanding of investment decisions. Concerning student subjects, Elliott et al. (2007) find that students are suitable subjects for tasks with low integrative complexity. Since our task does not demand complex information acquisition and integration, we argue that students behave similarly to real-world project managers in our experimental setting. Moreover, Graf-Vlachy (2019) shows that debiasing techniques usually are similarly effective for both students and managers if three criteria are met. First, EoC is present in student and manager populations (e.g., Mahlendorf 2015; Seybert 2010). Second, it can be expected that students and managers exhibit escalating

---

[10] To enable objective classification of continuing the poorly performing project as the wrong decision, our experimental instructions state that “projects with higher expected returns are preferred [by the firm]”. Since it is not likely in practice that firms have a clearly defined catalog of all project related mistakes, we provide a general picture of the given EMC and thus refer more vaguely to the project related error. Especially blame EMCs establish a general perception of bad outcomes being a mistake, irrespective of a justified prior decision-making leading to that result (Gold et al. 2014). Thus, we leave it to participants’ own perception on what they classify as the mistake.

[11] The institution where the study was conducted does not have a review board to provide ethics clearance. To conduct the experiment in an ethical manner, subjects were treated anonymously in accordance with data protection regulations. Furthermore, they were not exposed to specific risks, and they were not deceived by any means at any time.

tendencies due to the same psychological and social determinants. Third, de-escalating by highlighting negative information should have the same effect (increased salience of potential project failure) for students, as well as for managers.

After removing two participants, who erroneously indicated that the consultancy's assignment was to highlight positive information, our final sample consists of 126 participants. Participants' average age is 24.2 years, and 52.8% are female. There are no significant differences across conditions for age, gender, risk preference, bias knowledge, and the number of semesters studied (all p-values > 0.1).[12]

4 Results

4.1 Effectiveness of the Error Management Climate Manipulation

We assess the effectiveness of our EMC manipulation using six questions from our post-experimental questionnaire (Table 2). Questions 1 and 2 examine participants' assessment of how failure is dealt with in the organization, similar to van Dyck et al. (2005). They serve as an explicit manipulation check since both questions directly refer to statements displayed in the EMC manipulation. Questions 3 to 6 examine the impact our EMC manipulation had on participants' perception of psychological safety, which we measure with questions based on Edmondson (1999).

[Insert Table 2 about here]

To test if participants' perception of the firm's EMC differs between the open and blame EMC manipulation, we apply an explanatory factor analysis (EFA) to Questions 1 to 6. The EFA allows us to check whether our six EMC items converge to load on the latent construct of perceived EMC.[13] Cronbach's alpha (0.74) and Raykov's reliability coefficient (0.72) suggest

[12] Here and in the following, all p-values are reported as two-tailed.
[13] Allee et al. (2022) recommend using EFA to measure latent constructs and principal component analysis (PCA) to primarily reduce the number of variables. Thus, we apply EFA instead of PCA. However, both analyses yield

strong internal consistency of the six response items (Cho and Kim 2014; Raykov 1997). Based on the EFA, only the first factor has an Eigenvalue larger than one (2.19). All six items positively correlate with the first factor (factor loadings ranging between 0.47 and 0.72). The Kaiser-Meyer-Olkin measure of sampling adequacy indicates a common variance between the six questions (overall KMO = 0.65). Using the factor in a t-test, we find that participants in the open EMC treatment have significantly lower values of the factor (mean = -0.63) than participants in the blame EMC (mean = 0.67; t = -11.48, $p < 0.01$). Lastly, we apply individual two-sample t-tests to each of the six items to check for differences between both EMC treatments (Table 2). The responses to all six questions differ significantly depending on the EMC (all p-values < 0.012). Taken together, these findings suggest that our EMC manipulation was successful.

4.2 Test of the Research Questions and the Hypothesis

The descriptive statistics are presented in Table 3 Panel A. The dependent variable is the participants' recommendation to continue the poorly performing project, ranging from 0 (termination) to 100 (continuation). We find that continuation is similar in the open (54.05) versus the blame EMC (58.05) and higher when the devil's advocate is absent (64.98) than when present (47.80). The devil's advocate effect is slightly more pronounced in the open EMC treatment (63.22 vs. 45.15) than in the blame treatment (67.00 vs. 50.45).

[Insert Table 3 about here]

Our first research question (*RQ1*) examines whether the EMC affects EoC. ANOVA results are presented in Table 3 Panel B. The two independent variables are *devil's advocate* and *EMC*. The ANOVA results reveal no main effect for EMC (F = 0.61, p = 0.44). Thus, an

similar results (untabulated). Furthermore, we use EFA instead of confirmatory factor analysis (CFA) due to two reasons. First, there is no established measurement of EMC in prior literature, which makes EFA the recommended tool for measuring the construct (Bedford and Speklé 2018; Kelloway 1995). Second, a sample size with a minimum of 150 responses per item is recommended for a CFA with three or more indicators (Iacobucci 2010), which we do not exceed.

open EMC does not mitigate EoC. We further discuss this finding in the additional analysis section.

The ANOVA results display a significant main effect for the devil's advocate ($F = 8.83$, $p < 0.01$) and an insignificant interaction effect ($F = 0.02$, $p = 0.90$). To formally test *H1* and *RQ2*, we present simple effect tests following the ANOVA (Table 3 Panel C). *H1* predicts that the devil's advocate reduces EoC in an open EMC and *RQ2* examines whether the de-escalating effect of the devil's advocate also materializes in a blame EMC. The simple effects test results show that escalation is significantly reduced by the devil's advocate in an open culture (63.22 > 45.15, $t = -2.23$, $p = 0.027$) and in a blame culture (67.00 > 50.45, $t = -1.97$, $p = 0.051$). Thus, the results support *H1* and confirm *RQ2*.

4.3 Additional Analyses

In this section, we further substantiate our theory. First, we provide process evidence (Asay et al. 2021), which explains why the open EMC does not lead to lower escalation levels compared to the blame EMC. Second, we provide process evidence showing the drivers behind the de-escalating effect of the devil's advocate.

Process evidence: Why does an open EMC not lead to lower levels of EoC?

For *RQ1*, we do not find that EoC is lower in an open than in a blame EMC. We argue that the open EMC has a two-sided impact on escalating behavior. On the one hand, an open EMC might encourage exploration and risk-taking (Fischer et al. 2018; Edmondson 1999), and decision-makers do not expect any repercussions (e.g., a bad performance evaluation) in case of project failure. This could lead to experimentation and irrational project continuation. We measure participants' propensity of risk-taking and their concerns about future repercussions by asking them "*In my role as a leading manager at Kitchen World, I had concerns about taking risks*." and "*In case the management would evaluate my work performance, I am afraid I would get a bad evaluation."* (1 = fully disagree; 7 = completely agree).

On the other hand, an open EMC can induce the feeling of being safe for interpersonal risk-taking. Thereby, decision-makers are more likely to admit project failure of the initialized poorly performing project, resulting in less EoC (Cannon and Edmondson 2005, 2001). To approximate participants' concern about giving a wrong continuation recommendation, we asked them "*I was concerned about giving a wrong recommendation to the management*." (1 = fully disagree; 7 = completely agree).

[Insert Figure 2 about here]

Figure 2 shows process model 4 of Hayes (2018) with the three mediators. We use bootstrap resamples with replacements to compute path coefficients for the mediation model.[14] First, the open EMC has a significant negative impact on escalation ($\beta = -10.93$, $p = 0.06$). This negative effect is partially explained by participants' concerns about giving a wrong recommendation (*Mediator 3*), which is significantly decreased by the open EMC ($\beta = -1.24$, $p < 0.01$). Lower values of this item significantly decrease EoC ($\beta = 3.89$, $p = 0.052$).

Contrary to these de-escalating effects, the open EMC also enhances EoC. Participants in the open EMC are significantly less afraid of future repercussions (*Mediator 1*) ($\beta = -1.13$, $p < 0.01$) and are also significantly less afraid of taking risks (*Mediator 2*) ($\beta = -1.48$, $p < 0.01$). Decreases in both mediators lead to more escalation ($\beta = -3.54$, $p = 0.045$; $\beta = -5.23$, $p < 0.01$). Both mediators have a positive and significant indirect effect on EoC (BC CI repercussions: 4.15, 15.12; taking risks: 0.48, 7.70), whereas participants' decreased concerns about giving a wrong recommendation (*Mediator 3*) have a negative indirect effect on EoC (BC CI: -9.71, -0.77).

[14] We use bias-corrected bootstrap confidence intervals (BI CIs) to test for statistical significance of the respective indirect effect. BC CIs are most balanced in regard to type I error rates and statistical power (Chen and Fritz 2021; Tibbe and Montoya 2021). If zero does not appear within the bias-corrected CI, then the respective indirect effect has statistical significance at the 10% level.

Since the total indirect effect of the model is significantly positive (BC CI: 2.23, 12.08), the negative significant direct effect of the open EMC is attenuated, leading to an insignificant total effect of the open EMC on escalation (β = --4.00, $p < 0.50$). To summarize, the partial competitive mediation model shows a two-sided impact of the open EMC on EoC, explaining why decision-makers in the open EMC exhibit similar levels of EoC compared to the blame EMC.

Process evidence: What drives the successful de-escalating effect of the devil's advocate?

Using another mediation model, we analyze how the de-escalating effect of the devil's advocate is mediated by two well-established psychological drivers of EoC: 1) the goal-substitution effect (Conlon and Garland 1993) and 2) participants' focus on positive aspects (Caldwell and O'Reilly 1982). We argue that increased emphasis on goal completion –which increases EoC– can be mitigated by the devil's advocate. The goal-substitution effect is measured by asking participants whether "*It was important to me to complete the project*" (1 = fully disagree, 7 = completely agree). Furthermore, we argue that decision-makers' tendency to overweigh positive information is attenuated by the devil's advocate, which is measured by the item "*While giving my recommendation to the management board, I particularly focused on aspects of the project that went well*." (1 = fully disagree; 7 = completely agree).

[Insert Figure 3 about here]

Figure 3 shows process model 4 of Hayes (2018) with the two mediators. We use bootstrap resamples with replacements to compute path coefficients. Concerning the goal-substitution mediator, decision-makers put significantly less emphasis on completing the project (*Mediator A*) when the devil's advocate is present ($\beta = -0.56$, $p = 0.036$). Lower goal-substitution susceptibility significantly decreases project continuation ($\beta = 10.72$, $p < 0.01$). Concerning the second mediator, the devil's advocate significantly reduces the emphasis decision-makers put on positive aspects (*Mediator B*) ($\beta = -0.45$, $p = 0.082$), whereof lower values significantly reduce EoC (β = 5.06, $p < 0.01$). The direct effect of the devil's advocate significantly reduces EoC (β = -8.93, $p = 0.031$). Since the indirect effect of the goal-substitution effect (90% bias-corrected confidence interval: -11.04, -1.11) and the focus on positive information indirect effect are statistically significant (BC CI: -4.54, -0.09), both effects partially and complementary mediate the relation between the devil's advocate and EoC.

To summarize, the mediation model reveals that assigning someone to criticize a project sends a clear signal to the decision-maker that the goal of project completion should be reassessed, independent of the EMC. Furthermore, decision-makers' tendency to overweigh positive information in an EoC scenario is successfully attenuated by the devil's advocate in both EMCs.

5 Conclusion

Escalation of commitment (EoC) describes the tendency to overinvest resources in poorly performing projects (Staw 1976). Using an experiment, we investigate the impact of a devil's advocate –assigned to highlight negative information indicating project failure– in conjunction with a firm's organizational error management (EMC) on escalation.

First, we examine the impact of a blame (preventing errors) versus an open (learning from errors) EMC on escalating behavior. Prior research shows an ambiguous impact of firms' EMC on EoC (Fischer et al. 2018; Mahlendorf 2015; Barton et al. 1989). Our results do not show different escalation levels in both EMCs (*RQ1*). To substantiate this finding, process evidence reveals a two-sided impact of the open EMC on EoC (being able to admit project failure vs. increased risk-taking and not being concerned about repercussions), explaining the insignificant EMC main effect. Moreover, we find that the presence of the devil's advocate decreases participants' escalating behavior in both EMCs, supporting *H1* and confirming *RQ2*. Process evidence reveals that the devil's advocate reduces decision-makers' susceptibility to the goal-substitution effect and their tendency to overweigh positive information in both EMCs.

To further isolate the drivers behind the devil's advocate, future research could examine whether manipulating the amount of positive information or the amount of project progress impacts the effectiveness of the devil's advocate.

Based on our findings, we contribute to theory and practice in several ways. First, we show that implementing a devil's advocate with the simple assignment of providing highlighted negative information successfully mitigates EoC. Our de-escalation strategy has the advantage of not requiring a recommendation or interpretation of information, which could lead to false recommendations or misinterpretation of information (Behrens and Ernst 2014; Schwenk 1988). Our findings are particularly relevant for firms such as Google and Facebook that increasingly use debiasing techniques (Facebook 2022; Google Rework 2023). Second, we add to prior (accounting) EoC research. In contrast to other de-escalation tools (e.g., Brüggen and Luft 2016; Cheng et al. 2003), which come with potential caveats, we show that the devil’s advocate technique can effectively attenuate EoC by directly debiasing the project owner’s decision-making process, while keeping her expertise in the decision process. Thereby, we directly add to the findings of Loh et al. (2019), who examine how consultants behave in an EoC scenario. Third, we provide process evidence explaining why prior research finds mixed evidence concerning the impact of the EMC on escalation (Fischer et al. 2018; Mahlendorf 2015; Barton et al. 1989).

Our study is subject to several limitations, which offer avenues for future research. In our experimental design, we manipulate the EMC at two levels: open vs. blame. In practice, firms most likely exhibit elements of both EMC types. Even though several PEQ items prove the effectiveness of our EMC manipulation, using a fictitious experimental setting –without real-live social interactions– might not induce as strong effects on social EoC determinants as real corporate settings. Moreover, there could be a causality between a firm's EMC and the decision to implement a devil's advocate. Intuitively, it appears more likely that firms with an open EMC would implement a devil's advocate since they provide an environment where decision-makers are enabled to appreciate criticism. Nevertheless, due to a potential lack of trust from upper management in project owners, it is also plausible that firms with a blame EMC would implement the devil's advocate technique to uncover mistakes. Thus, future research could examine whether the propensity to implement debiasing tools (such as a devil's advocate) depends on firms' EMCs.

Furthermore, future research could manipulate the classification of the project information as private or public. This could affect decision-makers in the blame EMC, since project owners could have the perception that the management cannot objectively classify project continuation as a mistake. Contrarily, in the open EMC it could be expected that the transparency of information has no or less influence on EoC, since decision-makers do not expect any repercussions in case of project failure. Lastly, implementing a devil’s advocate could erroneously halt well-performing projects. To identify this potential effect, future research could manipulate whether decision-makers receive negative or positive project feedback on their initial decision.

Figure 1

Experimental Procedure

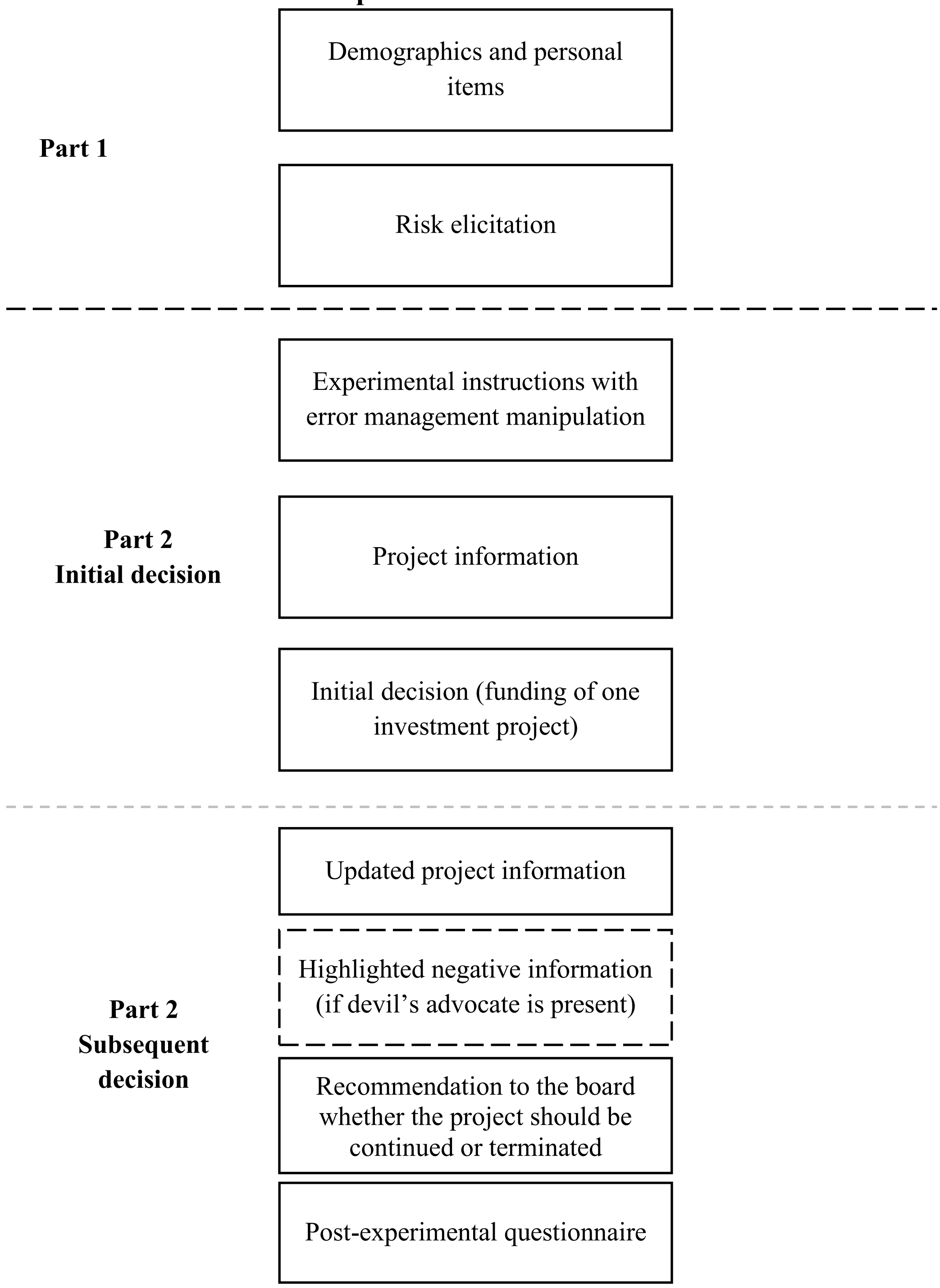
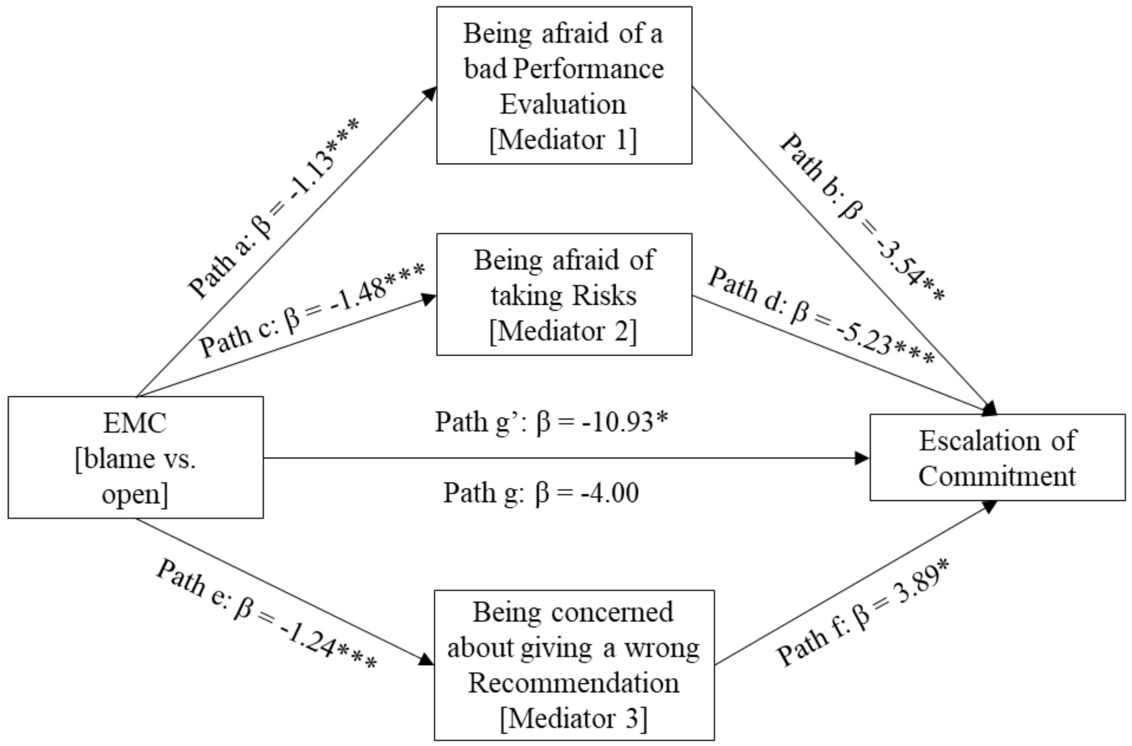


Figure 2
Mediation Model: The Effects of the Error Management Climate on Escalation of Commitment



90% bias-corrected CI for the "bad performance evaluation" indirect effect (a×b): **0.48, 7.70**
90% bias-corrected CI for the "taking risks" indirect effect (c×d): **4.15, 15.12**
90% bias-corrected CI for the "false recommendation" indirect effect (e×f): **-9.71, -0.77**

90% bias-corrected CI for the total indirect effect (a×b + c×d + e×f): **2.23, 12.08**

The figure displays mediation model 4 of Hayes (2018) with three mediators and their path coefficients. If zero does not appear within the bias-corrected confidence-interval, then the respective indirect effect has statistical significance at the 10% level (displayed in **bold**).

Path *a* reflects the direct effect of the independent variable Error Management Climate (EMC) blame (0) vs. open (1) on the first mediator "being afraid of a bad performance evaluation", which is derived from the post-experimental questionnaire item "In case the management would evaluate my work performance, I am afraid I would get a bad evaluation." (1 = fully disagree; 7 = completely agree).

Path *b* reflects the direct effect of the first mediator (*Mediator* 1) on the dependent variable Escalation of Commitment (EoC), which is approximated by participants' recommendation to continue a poorly performing project, measured on a 101-scale (0 = termination, 100 = continuation).

Path *c* reflects the direct effect of the independent variable EMC on the second mediator (*Mediator* 2) "being afraid of taking risks", which is derived from the post-experimental questionnaire item "In my role as a leading manager at Kitchen World, I had concerns about taking risks." (1 = fully disagree; 7 = completely agree).

Path *d* reflects the direct effect of the second mediator on the dependent variable EoC.

Path *e* reflects the direct effect of the independent variable on the third mediator (*Mediator* 3) "being concerned about giving a wrong recommendation", which is derived from the post-experimental questionnaire item "I was concerned about giving a wrong recommendation to the management." (1 = fully disagree; 7 = completely agree).

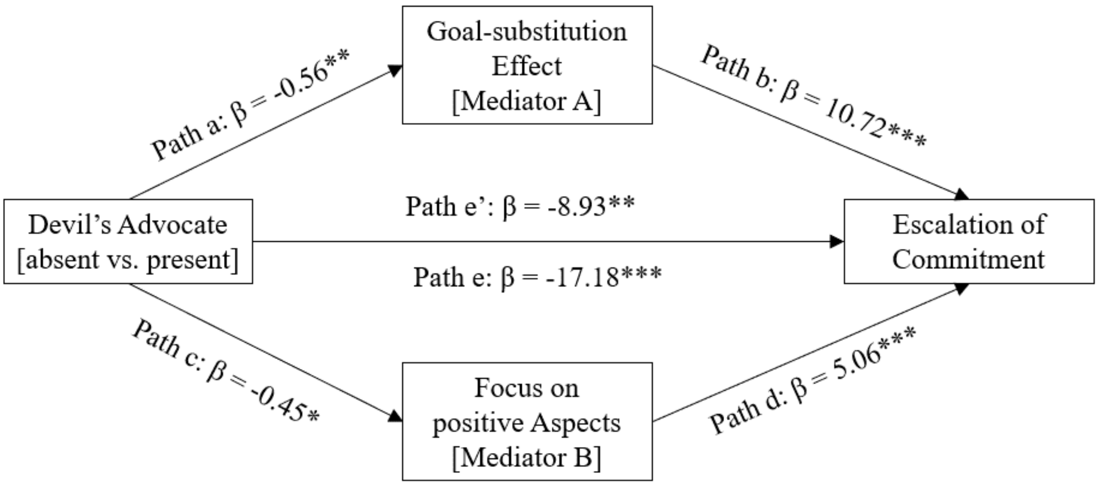
Path *f* reflects the direct effect of the third mediator on the dependent variable EoC.

Path *g'* reflects the direct effect of the independent variable EMC on EoC. Path *g* reflects the total effect (g' + (a×b + c×d + e×f)). If the total indirect effect is significant and path g' is not significant, then the model displays a full mediation. Since both the direct effect *g'* and the total indirect effect (a×b + c×d + e×f) are statistically significant but in opposite directions, the model displays a partial competitive mediation, rendering the total effect *g* insignificant.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Figure 3

Mediation Model: The Effects of the Devil's Advocate on Escalation of Commitment



90% bias-corrected CI for the "Goal-substitution" indirect effect (a×b): **-11.04, -1.11**
90% bias-corrected CI for the "Focus on positive Aspects" indirect effect (c×d): **-4.54, -0.09**

90% bias-corrected CI for the total indirect effect (a×b + c×d): **-13.29, -1.47**

The figure displays mediation model 4 of Hayes (2018) with two mediators and their path coefficients. If zero does not appear within the bias-corrected confidence-interval, then the respective indirect effect has statistical significance at the 10% level (displayed in **bold**).

Path *a* reflects the direct effect of the independent variable Devil's Advocate absent (0) vs. present (1) on the first mediator "goal-substitution effect" (*Mediator A*), which is derived from the post-experimental questionnaire item "It was important to me to complete the project." (1 = fully disagree; 7 = completely agree). Higher values indicate more susceptibility to the goal-substitution effect.

Path *b* reflects the direct effect of the first mediator on the dependent variable Escalation of Commitment (EoC), which is approximated by participants' recommendation to continue a poorly performing project, measured on a 101-scale (0 = termination, 100 = continuation).

Path *c* reflects the direct effect of the independent variable on the second mediator "Focus on positive Aspects" (*Mediator B),* which is derived from the post-experimental questionnaire item "While giving my recommendation to the management board, I particularly focused on aspects of the project that went well." (1 = fully disagree; 7 = completely agree). Higher values indicate more focus on positive aspects.

Path *d* reflects the direct effect of the second mediator on the dependent variable EoC.

Path *e'* reflects the direct effect of the IV on EoC. Path e reflects the total effect (*e'* + (a×b + c×d)). If the total indirect effect is significant and path e' is not significant, then the model displays a full mediation. Since both the direct effect e' and the total indirect effect (a×b + c×d) are statistically significant and have the same sign, the model displays a partial complimentary mediation.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Table 1

Error Management Climate Manipulation
(differences are displayed in **bold**)

Blame	Open
Panel A: Exemplary E-Mail that is shown in the experiment	
The board is convinced that wrong decisions are **always avoidable** and therefore attaches great importance to the fact that employees **do their utmost to avoid possible errors.**	The board is convinced that wrong decisions are **not always avoidable** and therefore attaches great importance to the fact that employees **accept and learn from wrong decisions.**
Pictured below is an exemplary e-mail in which the board evaluates a project of the two marketing managers, Ms. Becker and Mr. Schwarz: We discussed your poorly performing marketing campaign in detail during our last management meeting. This campaign incurred costs of 4 million Euros but did not yield any measurable effects.	
As you already know, we believe that such errors **are always avoidable**. **We see this failure as a sign of your incompetence. Consequently, the project result will** have a negative impact on your annual performance evaluation. Hence, **we will not be able to consider you** for a promotion next year. **We expect you to do your utmost to avoid such errors in the future.**	As you already know, we believe that such errors **are not always avoidable**. **We would also like to emphasize that the outcome of this individual project will not** have a negative impact on your annual performance evaluation. Hence, **you still have a good chance** for a promotion next year. **However, we encourage you to analyze the drivers leading to the project's results in order for you and your employees to learn from them for future projects.**
Panel B: Error Management Climate Quiz Items	
Q1: What is particularly important to Kitchen World's management?	
The board attaches great importance to the fact that employees **do their utmost to avoid possible errors.**	The board attaches great importance to the fact that employees **accept and learn from wrong decisions.**
Q2: How does the board deal with poorly performing projects?	
The board sees such project results as a sign of incompetence, which has a negative impact on the annual performance evaluation	**As long as such projects are used as learning opportunities, they have no** impact on the annual performance evaluation.

Table 2

Perception of the Error Management Climate

Question (7-point scale)	EMC Open mean (n = 65)	EMC Blame mean (n = 61)	t-stat (p-value)
1. I have the feeling that at Kitchen World, the first thing that is done is to look for someone to blame when mistakes are made.	2.43	5.25	9.05 (< 0.01)
2. At Kitchen World, mistakes are seen as an opportunity to learn from. (*1 = completely agree*)	2.17	5.66	13.18 (< 0.01)
3. I have doubts about taking responsibility for future Kitchen World projects.	2.82	3.54	2.56 (0.012)
4. I was concerned about giving a wrong recommendation to the management.	3.43	4.67	3.73 (< 0.01)
5. In case the management evaluates my work performance, I am afraid of getting a bad evaluation.	2.75	3.89	3.90 (< 0.01)
6. In my role as a leading manager at Kitchen World, I had concerns about taking risks.	3.09	4.57	4.41 (< 0.01)

To measure participants' perception of the Error Management Climate (EMC), participants answered six items in the post-experimental questionnaire. On a 7-point scale, participants indicated their agreement (1 = fully disagree; 7 = completely agree) to the abovementioned statements. We reverse the scale of Question 2 to align it with the other items, where higher values are generally more representative of a blame EMC.

Table 3

How Error Management Climate and the Devil's Advocate affect Recommendation of Project Continuation (Escalation of Commitment)

Panel A: Descriptive Statistics – Recommendation of Project Continuation [Mean (SD)]

	Devil's Advocate		
Error Management Climate	Absent	Present	
Blame	67.00 (30.20) n = 28	50.45 (36.30) n = 33	58.05 (34.39) n = 61
Open	63.22 (32.30) n = 32	45.15 (30.95) n = 33	54.05 (32.66) n = 65
	64.98 (31.13) n = 60	47.80 (33.58) n = 66	55.98 (33.44) n = 126

Panel B: ANOVA Model

Source of variation	df	MS	F-Statistic	p-value
Error Management Climate [*RQ1*]	1	646.89	0.61	0.44
Devil's Advocate	1	9391.20	8.83	< 0.01
Error Management Climate × Devil's Advocate	1	18.15	0.02	0.90
Error	122	1063.82		

Panel C: Simple Effects Tests

Comparisons	Contrast	t-stat	p-value
Effect of Devil's Advocate given an Open Error Management Climate [*H1*]	-18.07	-2.23	0.027
Effect of Devil's Advocate given a Blame Error Management Climate [*RQ2*]	-16.55	-1.97	0.051

The dependent variable Escalation of Commitment (EoC) is approximated by participants' recommendation to continue a poorly performing project, measured on a 101-scale (0 = termination; 100 = continuation). We manipulate Error Management Climate (EMC) at open (1) vs. blame (0) and the Devil's Advocate at absent (0) vs. present (1).

Appendix 1

Panel A: Assignment of the consultancy

In the meantime, the management board decided that an external consulting agency should always be involved in investment decisions.

The task assigned to the consulting agency is always stated as the following:

> Attached you find information for one of our current projects.
>
> In order to support our internal decision on whether the project should be continued or terminated, we ask you to identify aspects that explicitly speak against continuing the project.
>
> You do not have to give a specific recommendation on whether you consider this investment advantageous or disadvantageous.
>
> Instead, please only **highlight aspects that could potentially lead to the failure of the project**.
>
> ______________________________
>
> *Management Board Kitchen World*

The management has sent a corresponding assignment - based on the same information already provided to you (*project Smart Coffee, Update 30 June 2021*) - to the consulting agency.

Click the „*Continue*" button to proceed to the consulting agency's analysis, which highlights possible weaknesses of the project Smart Coffee.

Continue to the analysis

Panel B: Report of the consultancy

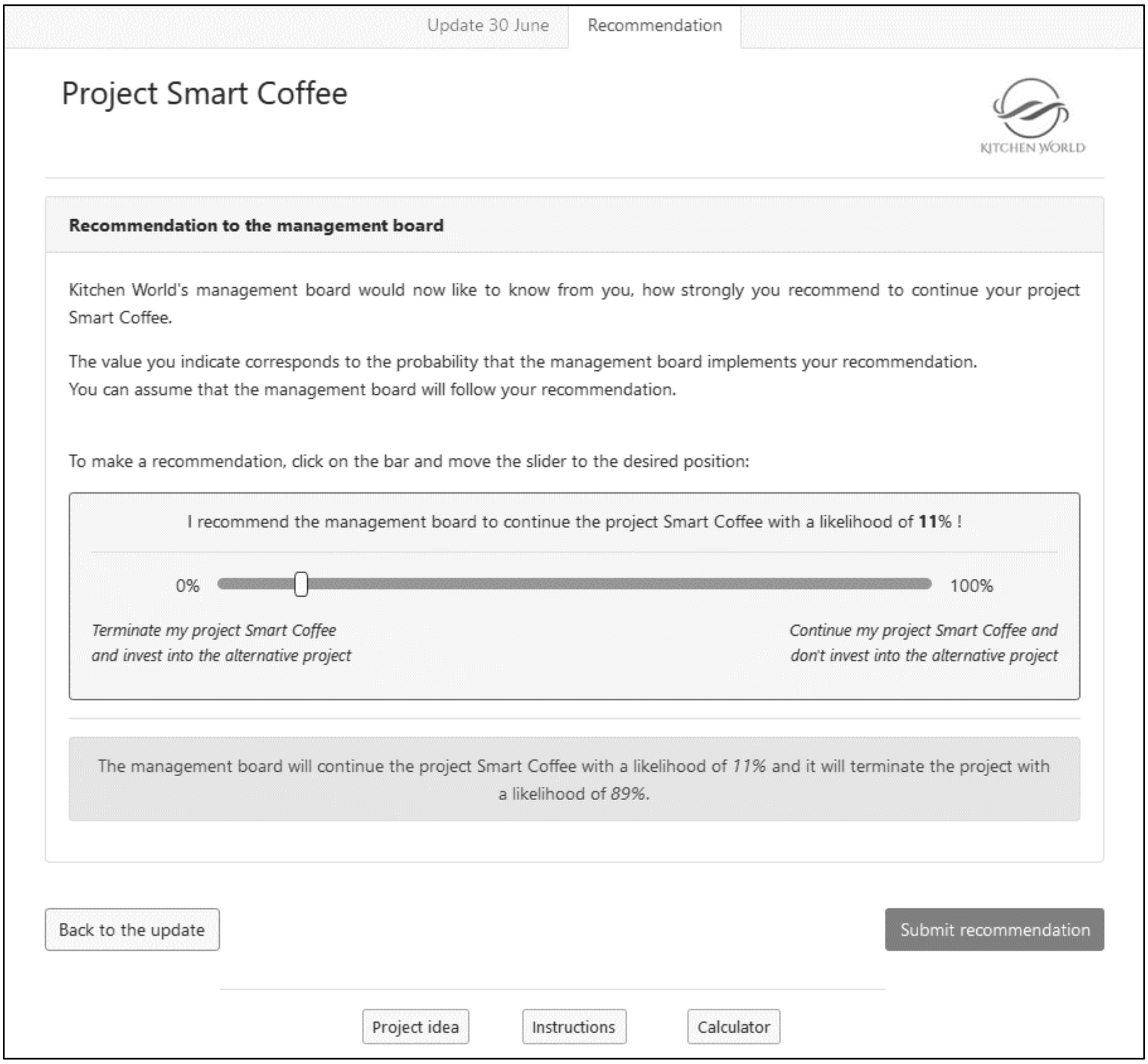
In the following you receive the analysis of the consulting agency:

> *Dear project management,*
>
> *since last year, you have been responsible for the Smart Coffee project, which our team thoroughly examined for potential weaknesses based on the given information.*
>
> *We were able to identify* ***two critical weaknesses in your project*** *that could eventually lead to* ***failure of your Smart Coffee project****:*
>
> 1. *The initial calculation of the project financials (before project launch) did not consider any serious competition. However, it currently seems that you will face competition, which will most likely diminish your market share.*
> 2. *There is currently a high degree of uncertainty whether the cooperation with Kitchen World's primary distribution partner can be used to sell Smart Coffee. If the cooperation fails, existing synergies that were included in the initial calculation of the project financials (before project launch) will be lost.*
>
> ______________________________
>
> *Senior Consultant*
>
> Consulting

Continue to the recommendation

Appendix 2

Recommendation to continue the project (dependent variable, Escalation of Commitment)


Update 30 June | Recommendation

Project Smart Coffee

KITCHEN WORLD

Recommendation to the management board

Kitchen World's management board would now like to know from you, how strongly you recommend to continue your project Smart Coffee.

The value you indicate corresponds to the probability that the management board implements your recommendation.
You can assume that the management board will follow your recommendation.

To make a recommendation, click on the bar and move the slider to the desired position:

I recommend the management board to continue the project Smart Coffee with a likelihood of **11%** !

0% 100%

Terminate my project Smart Coffee and invest into the alternative project

Continue my project Smart Coffee and don't invest into the alternative project

The management board will continue the project Smart Coffee with a likelihood of *11%* and it will terminate the project with a likelihood of *89%*.

Back to the update

Submit recommendation

Project idea | Instructions | Calculator

Appendix 3: Error management climate manipulation (blame treatment)

Information about your task +

Remuneration and experimental currency +

Information about the company −

A few years ago, *Kitchen World* got a new management board. The new board follows a different **strategical approach** by placing a stronger focus on household and kitchen appliances that incorporate elements of **modern technology.**

Besides the new strategical focus, **error management** is a crucial aspect within *Kitchen World's* daily business.

The board is convinced **that wrong decisions are always avoidable** and therefore attaches great importance to the fact **that employees do their utmost to avoid possible errors.**

Pictured below is an exemplary e-mail, in which the board evaluates a project of the two marketing managers Ms. Becker and Mr. Schwarz:

From: *<management.board@kitchen-world.de>*
Date: Thursday, 6 February 2020 03:10pm
To: *<Becker@marketing.kitchen-world.de> <Schwarz@marketing.kitchen-world.de>*
Subject: Evaluation: Marketing campaign "Product awareness in Eastern Europe"

Dear Ms. Becker,
Dear Mr. Schwarz,

we discussed your poorly performing marketing campaign in detail during our last management meeting. This campaign incurred costs of 4 million Euros but did not yield any measurable effects.

As you already know, we believe that such errors are always avoidable.

We see this failure as a sign of your incompetence.
Consequently, the project result will have a negative impact on your annual performance evaluation.

Hence, we will not be able to consider you for a promotion next year.

We expect you to do your utmost to avoid such errors in the future.

____

The Board

These instructions and a calculator are always available below the "Next" button during the quiz and the further experiment.

Continue to the quiz

References

Allee, K., Do, C., & Raymundo, F. G. (2022). Principal Component Analysis and Factor Analysis in Accounting Research. *Working Paper*.

Asay, H. S., Guggenmos, R., Kadous, K., Koonce, L., & Libby, R. (2021). Theory Testing and Process Evidence in Accounting Experiments. *The Accounting Review*.

Baer, M., & Frese, M. (2003). Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior, 24*, 45–68.

Baer, T., Heiligtag, S., & Samandari, H. (2017). McKinsey: The business logic in debiasing.

Barton, S., Duchon, D., & Dunegan, K. (1989). An Empirical Test of Staw and Ross's Prescriptions for the Management of Escalation of Commitment Behavior in Organizations. *Decision Sciences, 20*, 532–544.

Bedford, D. S., & Speklé, R. F. (2018). Construct Validity in Survey-Based Management Accounting and Control Research. *Journal of Management Accounting Research, 30*, 23–58.

Behrens, J., & Ernst, H. (2014). What Keeps Managers Away from a Losing Course of Action? Go/Stop Decisions in New Product Development. *Journal of Product Innovation Management, 31*, 361–374.

Booth, P., & Schulz, A. K.-D. (2004). The impact of an ethical environment on managers' project evaluation judgments under agency problem conditions. *Accounting, Organizations and Society, 29*, 473–488.

Brink, A. G., Gouldman, A., Rose, J. M., & Rotaru, K. (2020). Effects of superiors' compensation structures on psychophysiological responses and real earnings management decisions of subordinate managers. *Management Accounting Research, 48*, 100691.

Brockner, J., Rubin, J. Z., & Lang, E. (1981). Face-saving and entrapment. *Journal of Experimental Social Psychology, 17*, 68–79.

Brüggen, A., & Luft, J. L. (2016). Cost Estimates, Cost Overruns, and Project Continuation Decisions. *The Accounting Review, 91*, 793–810.

Caldwell, D. F., & O'Reilly, C. A. (1982). Responses to Failure: The Effects of Choice and Responsibility on Impression Management. *Academy of Management Journal, 25*, 121–136.

Cannon, M. D., & Edmondson, A. C. (2001). Confronting failure: antecedents and consequences of shared beliefs about failure in organizational work groups. *Journal of Organizational Behavior, 22*, 161–177.

Cannon, M. D., & Edmondson, A. C. (2005). Failing to Learn and Learning to Fail (Intelligently). *Long Range Planning, 38*, 299–319.

Charness, G., Haruvy, E., & Sonsino, D. (2007). Social distance and reciprocity: An Internet experiment. *Journal of Economic Behavior & Organization, 63*, 88–103.

Chen, D., & Fritz, M. S. (2021). Comparing Alternative Corrections for Bias in the Bias-Corrected Bootstrap Test of Mediation. *Evaluation & the health professions, 44*, 416–427.

Chen, D. L., Schonger, M., & Wickens, C. (2016). oTree—An open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance, 9*, 88–97.

Cheng, M. M., Schulz, A. K.-D., Luckett, P. F., & Booth, P. (2003). The Effects of Hurdle Rates on the Level of Escalation of Commitment in Capital Budgeting. *Behavioral Research in Accounting, 15*, 63–85.

Cho, E., & Kim, S. (2014). Cronbach's Coefficient Alpha: Well Known but Poorly Understood. *Organizational Research Methods, 18*, 207–230.

Conlon, D. E., & Garland, H. (1993). The Role of Project Completion Information in Resource Allocation Decisions. *Academy of Management Journal, 36*, 402–413.

Cosier, R. A. (1981). Dialectical Inquiry in Strategic Planning: A Case of Premature Acceptance. *Academy of Management Review, 6*, 643–648.

Denison, C. A. (2009). Real Options and Escalation of Commitment: A Behavioral Analysis of Capital Investment Decisions. *The Accounting Review, 84*(1), 133–155.

Edmondson, A. (1999). Psychological Safety and Learning Behavior in Work Teams. *Administrative Science Quarterly, 44*, 350.

Edmondson, A. (2002). Managing the risk of learning: Psychological safety in work teams. *Harvard Business School Working Paper, 2*.

Edmondson, A. C., & Lei, Z. (2014). Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 23–43.

Elliott, W. B., Hodge, F. D., Kennedy, J. J., & Pronk, M. (2007). Are M.B.A. Students a Good Proxy for Nonprofessional Investors? *The Accounting Review, 82*(1), 139–168.

Facebook. (2022). Managing Unconscious Bias. https://managingbias.fb.com/.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of personality and social psychology, 50*, 229–238.

Fehrenbacher, D. D., Kaplan, S. E., & Moulang, C. (2020). The role of accountability in reducing the impact of affective reactions on capital budgeting decisions. *Management Accounting Research, 47*, 100650.

Feld, B. (2011). Note To CEO's: Decisions Come From You, Not The Board. https://www.businessinsider.com/note-to-ceos-decisions-come-from-you-not-the-board-2011-7.

Fischer, S., Frese, M., Mertins, J. C., & Hardt-Gawron, J. V. (2018). The Role of Error Management Culture for Firm and Individual Innovativeness. *Applied Psychology, 67*, 428–453.

Frese, M., & Keith, N. (2015). Action errors, error management, and learning in organizations. *Annual review of psychology, 66*, 661–687.

Gold, A., Gronewold, U., & Salterio, S. E. (2014). Error Management in Audit Firms: Error Climate, Type, and Originator. *The Accounting Review, 89*, 303–330.

Google Rework. (2023). Unbiasing. https://rework.withgoogle.com/subjects/unbiasing/.

Graf-Vlachy, L. (2019). Like student like manager? Using student subjects in managerial debiasing research. *Review of Managerial Science, 13*, 347–376.

Greitemeyer, T., Schulz-Hardt, S., & Frey, D. (2009). The effects of authentic and contrived dissent on escalation of commitment in group decision making. *European Journal of Social Psychology, 39*, 639–647.

Gronewold, U., & Donle, M. (2011). Organizational Error Climate and Auditors' Predispositions toward Handling Errors. *Behavioral Research in Accounting, 23*, 69–92.

Gronewold, U., Gold, A., & Salterio, S. E. (2013). Reporting Self-Made Errors: The Impact of Organizational Error-Management Climate and Error Type. *Journal of Business Ethics, 117*, 189–208.

Günther, B., Heiligtag, S., & Webb, A. (2017). McKinsey Quarterly: A case study in combating bias.

Hambrick, D. C. (2007). Upper Echelons Theory: An Update. *Academy of Management Review, 32*, 334–343.

Harrell, A., & Harrison, P. (1994). An incentive to shirk, privately held information, and managers' project evaluation decisions. *Accounting, Organizations and Society, 19*, 569–577.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, London: The Guilford Press.

Heng, C.-S., Tan, B. C., & Wei, K.-K. (2003). De-escalation of commitment in software projects: Who matters? What matters? *Information & Management, 41*, 99–110.

Herbert, T. T., & Estes, R. W. (1977). Improving Executive Decisions by Formalizing Dissent: The Corporate Devil's Advocate. *Academy of Management Review, 2*, 662–667.

Highhouse, S., & Yüce, P. (1996). Perspectives, Perceptions, and Risk-Taking Behavior. *Organizational Behavior and Human Decision Processes, 65*, 159–167.

Hoffman, N., Huber, M., & Smith, M. (2018). An analytics approach to debiasing asset-management decisions. *McKinsey & Company*.

Hogarth, R. M. (1987). *Judgement and choice: The psychology of decision, 2nd ed* (Judgement and choice: The psychology of decision, 2nd ed). Oxford, England: John Wiley & Sons.

Iacobucci, D. (2010). Structural equations modeling: Fit Indices, sample size, and advanced topics. *Journal of Consumer Psychology, 20*, 90–98.

Ivancevich, J. M., Konopaske, R., & Matteson, M. T. (2011). *Organizational behavior and management* (9th ed.). New York, NY: McGraw-Hill Irwin.

Kadous, K., & Sedor, L. (2004). The Efficacy of Third-Party Consultation in Preventing Managerial Escalation of Commitment: The Role of Mental Representations. *Contemporary Accounting Research, 21*, 55–88.

Kahn, W. A. (1990). Psychological Conditions of Personal Engagement and Disengagement at Work. *Academy of Management Journal, 33*, 692–724.

Kaplan, R. S., & Mikes, A. (2012). Managing Risks: A New Framework. *Harvard Business Review*.

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior, 16*, 215–224.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels Issues in Theory Development, Data Collection, and Analysis. *Academy of Management Review, 19*, 195–229.

Kreilkamp, N., Schmidt, M., & Wöhrmann, A. (2021). Debiasing as a powerful management accounting tool? Evidence from German firms. *Journal of Accounting & Organizational Change, 17*, 555–582.

Loh, C.-Y., Coyte, R., & Cheng, M. M. (2019). Is a fresh pair of eyes always better? The effect of consultant type and assigned task purpose on communicating project escalation concerns. *Management Accounting Research, 43*, 1–14.

Mahlendorf, M. d. (2015). Allowance for failure: reducing dysfunctional behavior by innovating accountability practices. *Journal of Management & Governance, 19*, 655–686.

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review, 27*, 521–535.

O'Neill, O. A. (2009). Workplace Expression of Emotions and Escalation of Commitment. *Journal of Applied Social Psychology, 39*, 2396–2424.

Pan, G., & Pan, S. L. (2011). Transition to IS Project De-Escalation: An Exploration Into Management Executives' Influence Behaviors. *IEEE Transactions on Engineering Management, 58*, 109–123.

Plous, S. (1993). *The psychology of judgment and decision making*. Mcgraw-Hill Book Company.

Raykov, T. (1997). Estimation of Composite Reliability for Congeneric Measures. *Applied Psychological Measurement, 21*, 173–184.

Reason, J. (2000). Human error: models and management. *BMJ, 320*, 768–770.

Roetzel, P. G., Pedell, B., & Groninger, D. (2020). Information load in escalation situations: combustive agent or counteractive measure? *Journal of Business Economics, 90*, 757–786.

Schwenk, C. R. (1984). Devil's advocacy in managerial decision-making. *Journal of Management Studies, 21*, 153–168.

Schwenk, C. R. (1988). Effects of Devil's Advocacy on Escalating Commitment. *Human Relations, 41*, 769–782.

Schwenk, C. R., & Cosier, R. A. (1993). Effects of Consensus and Devil's Advocacy On Strategic Decision-Making. *Journal of Applied Social Psychology, 23*, 126–139.

Seybert, N. (2010). R&D Capitalization and Reputation-Driven Real Earnings Management (Partially Retracted). *The Accounting Review, 85*, 671–693.

Shepherd, D. A., Patzelt, H., & Berry, C. M. (2019). Why Didn't You Tell Me? Voicing Concerns Over Objective Information About a Project's Flaws. *Journal of Management, 45*, 1087–1113.

Simonson, I., & Staw, B. M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of Applied Psychology, 77*, 419–426.

Sleesman, D. J., Conlon, D. E., McNamara, G., & Miles, J. E. (2012). Cleaning Up the Big Muddy: A Meta-Analytic Review of the Determinants of Escalation of Commitment. *Academy of Management Journal, 55*, 541–562.

Sprinkle, G. B., Williamson, M. G., & Upton, D. R. (2008). The effort and risk-taking effects of budget-based contracts. Accounting, Organizations and Society, 33(4-5), 436-452. *Accounting, Organizations and Society, 33*, 436–452.

Staw, B. M. (1976). Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance, 16*, 27–44.

Staw, B. M., & Ross, J. (1987). Behavior in escalation situations: Antecedents, prototypes, and solutions. *Research in organizational behavior: an annual series of analytical essays and critical reviews, 9*, 39–78.

Tibbe, T., & Montoya, A. K. (2021). Correcting the Bias Correction for the Bootstrap Confidence Interval in Mediation Analysis. *Working Paper*.

Trigeorgis, L. (1996). *Real Options: Managerial Flexibility and Strategy in Resource Allocation* (Mit Press). MIT Press.

van Dyck, C., Frese, M., Baer, M., & Sonnentag, S. (2005). Organizational error management culture and its impact on performance: a two-study replication. *Journal of Applied Psychology, 90*, 1228–1240.

Wong, K. F. E., Yik, M., & Kwong, J. Y. Y. (2006). Understanding the emotional aspects of escalation of commitment: the role of negative affect. *The Journal of Applied psychology, 91*, 282–297.

Zhao, B., & Olivera, F. (2006). Error Reporting in Organizations. *Academy of Management Review, 31*, 1012–1030.

Zimmerman, J. L. (2019). *Accounting for decision making and control*. McGraw-Hill Education.

Zohar, D., & Luria, G. (2004). Climate as a social-cognitive construction of supervisory safety practices: scripts as proxy of behavior patterns. *Journal of Applied Psychology, 89*, 322–333.

Celebrating Failure – The Effects of Failure Awards on Risk-Taking and Escalation of Commitment

Authors	Rebecca Sabel, Hannes Gerstel, Arnt Wöhrmann

Abstract

Innovations and efficient resource allocation are essential for firm success. However, managers' 'fear of failure' prevents firms from achieving these goals. To counteract, firms started to grant Failure Awards. Failure Awards reward a promising idea or project that eventually fails. This study examines whether Failure Awards promote innovation through risk-taking and simultaneously reduce resource wastage by mitigating the commitment to a failing course of action (i.e., Escalation of Commitment (EoC)). We conduct an experiment in which we manipulate the type of Failure Award, i.e., whether it emphasizes promoting risk-taking or reducing EoC. In line with our predictions, we find that Failure Awards increase risk-taking, irrespective of the type. We further find that EoC is significantly reduced if the Failure Award focuses on promoting de-escalation. However, a de-escalation effect cannot be observed for Failure Awards that promote risk-taking. We show that psychological safety has a twofold effect that explains the results.

Publication Status

Preparing for submission

Contemporary Accounting Research (VHB-Jourqual 3: A)

Workshop/ Conference Presentations

2022 CAAA Annual Meeting (Saskatoon)

2022 AAA Annual Meeting (San Diego)

2022 ENEAR Conference and Doctoral Colloquium (Sevilla)

2023 EXRIMA Conference (Giessen)

Acknowledgements

We thank Jeremiah Bentley, Monika Schuhmacher, Willie Choi, Victor van Pelt and Todd Thornock for their helpful comments.

Author Contribution	Sabel	Gerstel	Wöhrmann
Relative Share	*60%*	*30%*	*10%*
Conceptualization	✓	✓	✓
Theory Development	✓	✓	
Methodology	✓	✓	
Data Acquisition	✓	✓	
Data Analyses	✓	✓	
Writing	✓	✓	✓

1 Introduction

Failure is ubiquitous in organizations and often unavoidable on the way to success. Yet, employees tend to exhibit a fear of failure. Survey results indicate that 40% of employees report a fear of failure and spend 20-40% of their time worrying about making mistakes (Brassey et al. 2019). The fear of failure comes with costs and has two major drawbacks to firms' competitiveness.

First, fear of failure prevents individuals from striving for innovations, which are crucial to firms’ growth, efficiency and productivity (Birkinshaw and Haas 2016a). This is because innovations are subject to a high degree of uncertainty and are closely linked to failure (Fischer et al. 2018). Thus, employees fear mistakes that may result from being innovative and avoid potential negative consequences, e.g., to their career or reputation, by exhibiting risk-averse behavior (García-Granero et al. 2015; Wu 2008; Zhou and George 2001). A survey by the Boston Consulting Group reports that 31% of the respondents identified a risk-averse culture as a key obstacle to innovation (Birkinshaw and Haas 2016b). Risk aversion leads to opportunity costs for (risk-neutral) shareholders if the manager refrains from investing in projects with the highest expected returns (Baysinger et al. 1991; Eisenhardt 1989; Wiseman and Gomez-Mejia 1998). Hence, it is in the firm’s interest to encourage employees to take risks to promote innovations and maximize firm returns.

Second, employees who fear failure are likely to fall prey to EoC (Johnson 2017). EoC is a cognitive bias by which a decision-maker takes an irrational decision to continue investing in a losing course of action, e.g., a poorly performing project, although withdrawal is economically preferred (Brockner 1992; Sleesman et al. 2012; Staw 1976). If employees fear failure, they are afraid to admit that it was a mistake to have started investing in the failing project in the first place. To prevent image losses, decision-makers keep investing in a failing project in the hope of turning it profitable again (Edmondson 2003; Sleesman et al. 2012). This is why

EoC is also known as "[o]ne of the most robust and costly decision errors" (Sleesman et al. 2012). Thus, there is a need to design management control systems that reduce escalation tendencies.

In practice, an increasing number of firms have started to grant *Failure Awards* to counteract the fear of failure and its negative impact on decision-making and firm performance (Johnson 2017; Morgan 2015).[1] Failure Awards are associated with no or only a symbolic financial reward but rely much more on "celebrating failure", e.g., by granting the award to employees during an official ceremony (Johnson 2017; Supercell 2021; TATA 2021). Astro Teller, the director of Google's R&D division 'Google X', provides an explanation for Google's decision to grant Failure Awards: "You must reward people for failing, […]. If not, they won't take risks and make breakthroughs. If you don't reward failure, people will hang on to a doomed idea for fear of the consequences. That wastes time […]." (Grossman 2014). This statement underlines the two goals of Failure Awards: (1) encourage risk-taking and thus innovations and (2) save resources through the early termination of failing projects (Johnson 2017; Leber 2016).

Whether Failure Awards can achieve these goals is still an open question. Until today, there is no empirical evidence on the effectiveness of Failure Awards. This is where our study contributes. Thus, we experimentally investigate whether Failure Awards can be used as a management accounting instrument to (1) promote *risk-taking* and (2) reduce *Escalation of Commitment (EoC)*.

[1] For instance, the marketing and communication agency Hill Holiday grants the "Epic Fail Award" (Proulx 2019). Proctor & Gamble introduced the "Heroic Failure Award" (Morgan 2015). Coca-Cola has an "Innovation Award" that celebrates projects that have failed (Clifford 2019). NASA grants the "Lean Forward; Fail Smart Award" (NASA 2021). Tata grants the "Dare to Try" award for failed projects (Waczek 2012). Supercell, a mobile game developer, opens a bottle of champagne for every failure (Supercell 2021). Google X rewards failure through applause (Leber 2016) and W.L. Gore, a manufacturing company, celebrates failing projects that were discontinued with beer and champagne (Deutschman 2004).

Anecdotal evidence shows that there are several types of Failure Awards. While the criteria to receive the Failure Award (i.e., (i) risk-taking, (ii) failure, (iii) deliberate discontinuation) are kept constant, the types differ on whether the award emphasizes the risk-taking (*innovation*) or the timely termination of a failing course of action (*discontinuation*) aspect more strongly.[2] Given this continuum, our study uses the two endpoints to examine whether the effect of Failure Awards on risk-taking and EoC differs by which goal is emphasized more strongly. Hence, we distinguish between an *innovation* and *discontinuation type* of Failure Award.

We argue that –irrespective of the type– Failure Awards reduce fear of failure by inducing psychological safety, i.e., the secured feeling that enables interpersonal risk-taking (Edmondson 1999). First, decision-makers are not afraid of negative consequences to their image or career when they must admit failure, which is why they are more willing to start risky projects. Second, decision-makers feel safe to experiment and take risks. Thus, they are actively encouraged to start risky projects and are also rewarded in case of failure. Consequently, we predict in Hypothesis 1a (Hypothesis 1b) that risk-taking is higher when discontinuation-type (innovation-type) Failure Awards are granted than when Failure Awards are absent.

For our second hypothesis, we again rely on psychological safety. On the one hand, decision-makers feel safe to accept failure and are more willing to discontinue a failing project, as they do not anticipate negative consequences to their career and self-image. Hence, Failure Awards are likely to reduce EoC. On the other hand, it is questionable whether this deescalating effect can still be observed when an innovation instead of a discontinuation type is used. If the

[2] For example, Hill Holliday introduced the "Epic Fail Award" to "[…] cultivate the kind of guts and appetite for risk-taking that's required of true innovators." (Proulx 2019). Similarly, Proctor & Gamble give out the "Heroic Failure Award" for taking the greatest 'intelligent' risk (Anthony 2020). W.L. Gore, on the other hand, celebrates failure with beer and champagne when "a project doesn't work out and the team kills it" (Deutschman 2004), thus emphasizing the early termination of a failing project. The "Innovation Award" from Coca-Cola stresses the importance of "killing zombies", i.e., killing products that do not work, which emphasizes the need for de-escalation (Clifford 2019).

firm highlights that the award is intended to increase risk-taking and thus innovation, employees are encouraged to experiment and may hold on to a failing project as they take the risk to bet on the small chance to turn the project profitable. Hence, our second hypothesis predicts that EoC is reduced when Failure Awards highlight discontinuation (H2). Due to a potential two-sided effect of Failure Awards that highlight innovation, we pose a research question of whether they are also effective in reducing EoC (RQ2).

To test our predictions, we employ a 2×1+1 between-subjects experimental design. Participants in the Failure Award absent treatment do not receive a Failure Award. Nested within the Failure Award present condition, we manipulate the *type* of the award at two levels (*innovation* vs. *discontinuation*). The type is derived from practical examples and either emphasizes the importance of taking risks and innovating or stopping wasting resources in failing projects.

In the experiment, participants must decide whether to invest in a project with low-risk (and lower expected returns) or in a project with high-risk (and higher expected returns). Risk-taking, our first dependent variable, is measured based on the selected project. Similar to Seybert (2010), Brink et al. (2020) and Denison (2009), participants learn that future returns are lower than expected, indicating project failure. Participants are asked to recommend to the management board whether the project should be continued. This recommendation is our second dependent variable that captures EoC.

As predicted by H1a and H1b, we find that Failure Awards increase risk-taking irrespective of their type. Further, we find that EoC decreases when discontinuation-type Failure Awards are used (H2). Regarding our research question (RQ2), we find that innovation-type Failure Awards do not reduce EoC. Additional analyses reveal that psychological safety has a twofold effect. The factor of "feeling safe to admit mistakes" reduces EoC. However, the effect of the second factor of "feeling safe to experiment and take risks" fosters EoC. This effect is exclusively triggered when the innovation type is introduced and sheds light on our research question. Further, the perceived risk, e.g., to one's image or career, is low in both Failure Award types since the mere granting of Failure Awards sufficiently reduces the risk individuals bear, which leads to increased risk-taking.

To the best of our knowledge, this study is the first to examine Failure Awards empirically, contributing to both practice and accounting theory. From a practical perspective, this paper draws attention to the concept of rewarding failures, which has gained increasing popularity in practice (e.g., Google X's "Failure Award", P&G's "Heroic Failure Award" and TATA's "Dare to Try Award" (Morgan 2015)) but has been neglected in research. Prior research that builds on similar constructs, e.g., implementing a culture that tolerates failure, finds that such a culture does not always effectively reduce EoC. Whereas some papers find a decrease in EoC through creating a failure tolerating culture (e.g., Simonson and Staw 1992), others find an increase in EoC (e.g., Barton et al., 1989). A potential explanation for these controversial findings is that project termination is not incentivized. Failure Awards, however, do not solely signal that failures are tolerated but reward them. Hence, through the emphasis that Failure Awards are granted only when failing projects are terminated, they are able to effectively decrease EoC. Furthermore, previous studies only examine the effect of failure tolerance on EoC and not the simultaneous effect on EoC and risk-taking.

Second, we provide important implications for implementing and designing (nonmonetary) incentive schemes in the form of Failure Awards. Specifically, we show that to promote innovations and risk-taking, the type of Failure Award is irrelevant. However, the type becomes crucial in regard to EoC. Our results show that only Failure Awards emphasizing project termination significantly reduce EoC. This is important as the majority of firms rather focus on promoting innovations and risk-taking and often neglect to emphasize de-escalation (examples are the "Epic Fail Award" by Hill Holiday (Proulx 2019), the "Heroic Failure Award" by Proctor & Gamble (Anthony 2020), or the “Lean Forward; Fail Smart Award” (NASA 2021)). Consequently, we inform firms to pay close attention to whether the type of Failure Award fits their goals.

From a theory perspective, we expand the intangible reward and social recognition literature that build on symbolic rewards and have received little attention so far. Especially, we add a new perspective by not restricting recognition to “best performance” and successful outcomes but instead reward failure. Moreover, we add to the literature stream that investigates risk-taking behavior. We show that Failure Awards can overcome the prevailing risk aversion of decision-makers. Furthermore, we contribute to the accounting phenomenon of EoC research (Cheng et al. 2003; Mahlendorf 2015). We identify discontinuation-type Failure Awards as a new and cost-efficient debiasing tool that reduces EoC. Failure Awards require only a low input of resources compared to other de-escalation strategies, e.g., hiring a third-party expert (Behrens and Ernst 2014).

Lastly, we shed light on the psychological mechanism that reduces the fear of failure through psychological safety (Frazier et al. 2017). Hence, we provide evidence that psychological safety has a twofold effect. First, we show that the feeling of being safe to admit failure reduces EoC. Second, we show that the (often neglected) effect of psychological safety –the feeling of being safe to experiment and take risks– is likely to encourage escalation behavior.

2 Background and Hypotheses Development

2.1 Failure Awards, Error Management Climate, and Psychological Safety

Generally speaking, failure can be seen as the lack of success, the outcome of “bad luck” or the inability to achieve a desired outcome or goal. Organizations typically strive for high performance by installing management processes based on predictability and efficiency, leaving limited to no room for failure (Birkinshaw and Haas 2016a; van Dyck et al. 2005). In addition, firms often link a decision-maker’s salary (e.g., bonuses) and reputation to error-free decisions and successful outcomes. This is why employees develop a "fear of failure" that can be defined as the "[…] disposition to avoid failure and/or the capacity for experiencing shame and humiliation as a consequence of failure." (Atkinson 1957).

To counteract the fear of failure, firms have recently started to grant Failure Awards (Johnson 2017; Kuvalekar and Ravi 2019). Such awards are granted to employees who have shown their willingness to innovate and take risks but ended up failing. Failure Awards honor employees often in the form of award ceremonies to express the company's appreciation and are usually associated with no or only a symbolic financial reward. Thus, Failure Awards are part of the intangible incentive system of a firm.[3]

Failure Awards serve two goals at once: (1) encouraging innovation by making it safe to take risks and (2) saving resources by making it safe to admit failures and abandon failing projects (Johnson 2017; Leber 2016; Morgan 2015). Firms use different types of Failure Awards to emphasize one goal more than the other. For instance, NASA, America's civil space program, describes its '*Lean Forward; Fail Smart Award*' as "[…] an award designed to encourage, recognize, and celebrate the spirit that propels individuals to take the risk to innovate, unfortunately failing to reach the desired outcome […]" (NASA 2021). Thus, this type emphasizes risk-taking (*innovation type*); however, to receive the award, proper handling of the failure (i.e., the deliberate decision to terminate a failing project) is also required. In contrast, Coca-Cola's *Innovation Award* stresses the importance of "killing zombies", i.e., killing products or projects that do not work (Clifford 2019). While the original project idea must be innovative, Coca-Cola highlights the goal of discontinuing failing projects (*discontinuation type)*.

[3] Given the practical examples, Failure Awards usually imply applause, trophies or award ceremonies (Stewart 2015). However, Google X states that "Google's rewards aren't big enough to encourage people to pull the plugs on their projects indiscriminately, but the rewards aren't small change either." (Johnson 2017). This suggests that some awards might also include a very small monetary reward.

Practical examples show that firms use their own criteria for awarding Failure Awards (e.g., Google X (Johnson 2017)). Based on these examples, we identify and define the following criteria: Employees are eligible to receive a Failure Award if they (a) took the risk of initiating an innovative project but (b) the project failed and thus (c) the employee deliberately decided to end the failing project in time. While the type, i.e., whether (a) the risk-taking or (c) the de-escalation aspect is emphasized, might change, the three criteria must always be met simultaneously. Consequently, employees do not qualify for a Failure Award when their innovative project fails due to external factors without deliberate and timely termination.

Failure Awards can be seen as a specific instrument of an open error management culture (EMC).[4] EMC is defined as a set of shared beliefs, norms, and common practices regarding how errors are dealt with in an organization (van Dyck et al. 2005). In an open EMC, failures are seen as acceptable outcomes and opportunities to learn from (Fischer et al. 2018; Gold et al. 2014; van Dyck et al. 2005). Failure Awards are likely to create an open EMC. Yet, they go a step further and do not just communicate that failures are tolerated but actively reward them. It can be observed that simply communicating and encouraging an open EMC is not always sufficient (Freiberg 2011).

For example, Barton et al. (1989) find in their experiment that implementing an open EMC, by telling participants that their initial decision to invest in a project demonstrated good judgment even though the project threatens to fail, increases investments in the failing project. One potential explanation is that participants are more likely to accept failure through decreased fear of failure and thus hold on to failing projects. However, Failure Awards incentivize project termination as they are explicitly granted when a failing project is deliberately terminated. Hence, Failure Awards are likely to overcome the limitations of an open EMC.

[4] Some studies use the term ‘error management climate’, but as culture and climate are inherently difficult to differentiate, both concepts are treated interchangeably in this study.

Similarly to an open EMC, Failure Awards induce psychological safety (Baer and Frese 2003; Cannon and Edmondson 2005; Edmondson and Lei 2014; James et al. 1977). In a high psychological safe environment, individuals feel safe to take interpersonal risks, as they do not fear negative consequences to their status or career (Edmondson 1999; Kahn 1990). According to Edmondson 2003, individuals make their decision whether to take a potential action by assessing the interpersonal risk associated with that action (e.g., the risk of being seen as incompetent). Failure Awards induce psychological safety by credibly signaling –through granting an award– that (project) failure does not result in any negative consequences to one's image or career. Failure Awards acknowledge the courage of engaging in promising but risky endeavors and demonstrate appreciation by granting an award during a ceremony. Consequently, employees feel psychologically safe and do not fear failure as there is no interpersonal risk associated with failure.

However, psychological safety has a twofold effect. First, it encourages employees to admit mistakes. Hence, individuals do not perceive their (non-monetary) wealth to be at risk if a project turns out to be a mistake, as neither their employment nor their reputation are at risk. The reduced perceived risk of engaging in interpersonal risk-taking increases the willingness to take risks (Keil et al. 2000; Sitkin and Weingart 1995; Wong 2005). In this vein, Palanski and Vogelgesang (2011) find a positive relationship between employees' perceptions of psychological safety and risk-taking. Second, psychological safety encourages experimentation and risk-taking by inducing the feeling of being protected from any negative consequences in case of failure (Baer and Frese 2003; Newman et al. 2017). To support this argument, Fischer et al. (2018) find that high psychological safety increases innovativeness, as exploration and experimentation behavior is encouraged. Hence, a psychologically safe environment might evoke the perception of being safe to experiment and take risks (Edmondson 1999). Moreover, Failure Awards do not only send a clear signal to decision-makers that adequate risk-taking is valued, but require adequate risk-taking as a precondition for receiving the award. Hence, firms signal that risk-taking is desired.

2.2 The Effect of Failure Awards on Risk-Taking Behavior

In this section, we focus on the effect of Failure Awards on risk-taking. While risk-taking can be easily described as the choice of a risky decision (Barki et al. 1993), the definition of risk appears more complex. Yet, the various definitions of risk exhibit two similarities: (1) the *probability* that an undesirable outcome occurs and (2) the *consequences* resulting from it (e.g., losses or decreased returns) (Barki et al. 1993; Highhouse and Yüce 1996; Sitkin and Pablo 1992). Thus, risk is expressed through the variance of the expected decision outcomes.

Agency theory assumes that agents (i.e., employees) are risk-averse (Eisenhardt 1989; Wiseman and Gomez-Mejia 1998). One driver of risk aversion is the fear of trying something new or uncertain (e.g., starting a new and risky project) (Lerner and Keltner 2001; Tsai and Young 2010). The sources of fear are numerous and versatile and are linked to potential failure. The most predominant ones are the fear of decreasing one's personal wealth (e.g., loss of bonus payments) (Wiseman & Gomez-Mejia, 1998); the fear of negative consequences (e.g., losing the chance for promotion) (Edmondson 1999); and the fear of failing (i.e., risking one's self-image) (Edmondson 2003; Zhou and George 2001).

Fear is seen as a perceived interpersonal risk when making decisions (Edmondson 2003; Wiseman and Gomez-Mejia 1998). Consistent with the Behavioral Agency Model (BAM), *risk-bearing*, i.e., the perceived wealth at risk, is a crucial factor that causes risk aversion (Wiseman and Gomez-Mejia 1998). Based on this model, individuals are less likely to engage in risk-taking the greater they perceive their wealth to be at risk. In the case of failure, employees are likely to face several negative consequences that put their wealth at risk. On the one hand, monetary consequences may materialize if an employee's compensation is based on the success of a project's outcome. On the other hand, indirect monetary and non-monetary consequences may also incur. The first results if future career and promotion prospects are harmed, while the latter results from reputation and image losses of the decision-maker (Hirshleifer 1993). As greater risks also imply a greater possibility of failure, the propensity to initiate risky projects decreases for risk-averse decision-makers. Supporting this argument, Sitkin and Weingart (1995) find that the degree to which decision-makers engage in risk-taking is negatively linked to their level of perceived risk inherent in a situation.

Although Failure Awards cannot compensate for financial declines in performance-based salary, they are likely to reduce the indirect consequences. As described earlier, Failure Awards induce psychological safety. We argue that the type of Failure Award influences which factor of psychological safety is strengthened. If the Failure Award type focuses on saving resources by emphasizing to stop failing projects (discontinuation type), decision-makers will feel *safe to admit failures (PS-I)*. However, if the Failure Award type focuses on innovation and experimentation (innovation type), decision-makers additionally perceive the environment as *safe to take risks and experiment (PS-II)*. We predict that both psychological safety factors, and thus Failure Award types, have a positive effect on risk-taking.

Based on BAM, the main effect that increases risk-taking and mitigates the driver of risk aversion results from reducing the perceived wealth at risk, which is given in both types. In the discontinuation type, individuals feel safe to admit mistakes. Thus, they do not perceive their wealth at risk in case of failure and are more likely to initiate risky projects. The effect of feeling safe to experiment and take risks in the innovation type directly reduces an individual's risk perception and thus encourages risk-taking. Hence, both types support the risk inducing effect of psychological safety.

Following this reasoning, we posit Hypothesis 1a and 1b as follows:

H1a: **Risk-taking is higher when discontinuation-type Failure Awards are granted than when Failure Awards are not granted.**

H1b: **Risk-taking is higher when innovation-type Failure Awards are granted than when Failure Awards are not granted.**

We do not make a prediction on whether there is a differential effect between both Failure Award types based on two reasons. First, the mandatory condition to receive a Failure Award is the engagement in risk-taking.[5] Hence, irrespective of the type, decision-makers can receive the firm's acknowledgment through a Failure Award only by choosing a risky option. Second, at the time decision-makers have to decide whether to start a risky endeavor (e.g., a risky project) or not, they are not yet emotionally attached to it, which allows a more objective evaluation of the decision. Studies have shown that emotions, specifically emotional resonance, influence the effectiveness of how objectives are worded or framed (Druckman and McDermott 2008).

In our setting, this is transferable to the Failure Award types, which differ in their wording and thus on the goal they emphasize. Emotional resonance is likely to occur through the identification with the subject of decision, e.g., a project (Giorgi 2017). Such identification is less likely to occur during the first stage of initiating a (risky) project (Hennig et al. 2023). Due to the emotional distance, the choice of engaging in risk-taking is based on objective evaluation, e.g., the project's expected value, and whether one qualifies to receive a Failure Award. As both factors depend on the project choice and not the type of Failure Award, we do not make a prediction on the difference between the two types.

---

[5] Most practical examples primarily focus on promoting risk-taking through their implemented Failure Award. For example the *Lean Forward; Fail Smart Award* from NASA is granted in order to " […] encourage […] the spirit that propels individuals to take the risk to innovate […]" (NASA 2021) and Proctor & Gamble grant the "Heroic Failure Award" to employees who took the greatest 'intelligent' risk (Anthony 2020). Therefore, we assume that decision-makers have to take at least some amount of risk to be eligible for a Failure Award.

2.3 The Effect of Failure Awards on Escalation of Commitment

After selecting and initiating a project, managers often stay committed to them. Hence, we derive our second hypothesis and research question on how Failure Awards affect Escalation of Commitment. Staw and Ross (1987) classify determinants of EoC in project-, psychological-, social- and structural-determinants. We focus on psychological and social determinants since Failure Awards appear to have the largest impact on them.

The psychological determinants are explained by self-justification theory (Festinger 1957; Sleesman et al. 2012). According to this theory, decision-makers feel the need to justify their decision to have started a poorly performing project (Brockner 1992; Sleesman et al. 2012). Incurred sunk costs facilitate self-justification pressures since decision-makers try to avoid being seen as resource wasters (Arkes and Blumer 1985).[6] Hence, decision-makers escalate to avoid psychological costs in case of (project) failure. The social determinants of EoC imply that other parties, such as evaluators or rivals, affect the decision-maker through the image they have of her (Sleesman et al. 2012; Staw and Ross 1989). According to self-presentation theory (Goffman 1959), people aim to manage the impressions others have of them. Therefore, they are reluctant to engage in behaviors that could threaten their image, e.g., admitting a failure by withdrawing from the initial course of action (Edmondson 2003; Sleesman et al. 2012). Thus, decision-makers stay committed to their initial decision.

To conclude, decision-makers engage in escalation behavior due to the fear of consequences they face when admitting a failure (e.g., reputation losses or career threats) (Mahlendorf 2015; Sleesman et al. 2012). These consequences increase self-justification pressure and impression management concerns. We argue that Failure Awards can mitigate both factors by inducing psychological safety (PS-I). Failure Awards induce psychological safety by signaling

[6] Sunk costs is one of several drivers of EoC. We elaborate more on sunk costs in the additional analysis.

that project failure does not indicate poor performance of the decision-maker. Hence, if an employee perceives high psychological safety and starts a project that ends up failing, she will not fear negative consequences for her image or career. Therefore, Failure Awards reduce the fear of either having to justify why a failing project was initially started (self-justification pressure) or being seen as an incompetent decision-maker (impression management concerns).

Supporting our argument, Simonson and Staw (1992) find that self-justification pressure can be decreased by telling participants that their previous decisions, which resulted in negative outcomes, are not an indicator of their intelligence. Similarly, Heng et al. (2003) find that assuring decision-makers that a superior's opinion about them will not be affected by the project's outcome, reduces EoC. Lastly, Mahlendorf (2015) provides evidence that organizational allowance for failure reduces managers' perceived threat of project failure, which reduces EoC. Thus, high psychological safety reduces the decision-maker’s reluctance to terminate a failing project.

However, psychological safety can also have the opposite effect. This occurs if the feeling of being protected from negative consequences encourages experimentation and risk-taking (PS-II) (Baer and Frese 2003; Newman et al. 2017). As discussed earlier, several studies have found a positive relationship between psychological safety and risk-taking or innovation (Fischer et al. 2018; Frazier et al. 2017; Kark and Carmeli 2009; Palanski and Vogelgesang 2011; Zhou and Pan 2015). Due to feeling safe to experiment, the second effect of psychological safety might encourage decision-makers to take the risk of committing to their failing project. Again, we argue that the type of Failure Award influences which psychological safety factor predominates. By implementing discontinuation-type Failure Awards, decision-makers feel safe to admit mistakes, i.e., admit that their project is failing. Further, they receive a clear signal that inadequate risk-taking (i.e., wasting resources) is not encouraged. Thereby, the de-escalating effect of high psychological safety (PS-I) is emphasized.

However, if the Failure Award type encourages innovation and risk-taking, decision-makers additionally perceive the environment as safe to take (inadequate) risks (PS-II). Thereby, they might bet on the small chance of turning the project profitable by project continuation. Thus, for innovation-type Failure Awards, it is questionable whether the de-escalating effect of induced high psychological safety (PS-I) is offset by the risk-encouraging effect of high psychological safety (PS-II). In contrast to discontinuation-type Failure Awards, the innovation type does not clearly signal that (inadequate) risk-taking is discouraged. Consequently, we have no directional prediction of whether innovation-type Failure Awards reduce EoC. We formally state H2 as a directional hypothesis and RQ2 as a research question:

H2: **Discontinuation-type Failure Awards reduce Escalation of Commitment.**

RQ2: **Do innovation-type Failure Awards reduce Escalation of Commitment?**

3 Research Design

3.1 Experimental Design and Procedure

To test our predictions and answer our research question, we employ a 2×1+1 between-subjects experimental design. We manipulate the type of Failure Award at two levels: *innovation type* and *discontinuation type.* Furthermore, a Failure Award absent treatment (control group) is employed in which Failure Awards are not provided. The experiment was conducted online on Amazon Mechanical Turk (MTurk) and programmed using oTree (Chen et al. 2016).

The experimental procedure is summarized in Figure 1. The experiment consists of five parts: (1) an eligibility check, (2) a lottery task, (3) a risk task, (4) an EoC task and (5) a post-experimental questionnaire (PEQ).[7] Our main task requires knowledge of the expected value

[7] A pretest revealed that the Failure Award was perceived by some as undesirable due to the negative connotation of 'failure'. In an experimental setting, the possibilities to convincingly present the Failure Award that mitigates these concerns are – compared to a firm setting – limited. The CIO of Hill Holliday who grants the Epic Fail Award states that "[d]espite its awful-sounding name, this award has become something that Hill Holliday employees strive to win." (Proulx 2019). Accordingly, we made two adjustments. We implemented a cheerful video

calculation, as we provide expected project cash inflows and probabilities. Thus, only participants who calculated the correct expected value of a prize wheel during the (1) eligibility check could participate. Next, participants complete a (2) lottery task. The lottery task is an established risk-elicitation instrument that measures ex-ante risk preferences (Sprinkle et al. 2008). Similar to Sprinkle et al. (2008), 15 scenarios are presented. Each scenario consists of a safe payment of $0.75 and a lottery that pays either $1.50 with a probability of p or $0 with a probability of (1-p). The probability p decreases from 85% (scenario 1) to 15% (scenario 15) in 5% increments. Participants indicate in which scenario they would like to switch from the lottery to the safe payment or if they always want to participate in the lottery.

We refer to participants choosing scenario eight as risk-neutral participants, as the expected value of the lottery equals the safe payment in this scenario. Consequently, participants switching from the lottery to the safe payment before (after) scenario eight are classified as risk-averse (risk-seeking).

[Insert Figure 1 about here]

After completing the lottery task, participants read the instructions. Participants learn that they will act as project managers at a fictitious company, 'CleverClean'. In the compensation description, participants are informed that one dollar equals 20,000 lira, the experimental currency. The compensation consists of (i) a fixed payment ($1.00), (ii) a payment from the lottery task, and (iii) a variable payment from the main task.[8] The variable payment is based on the final project account balance, which is determined by participants' decisions.[9] Subjects are

of an award ceremony and renamed the award ‘Courage Award’. Practical examples show that Failure Awards have numerous names. Some have negative connotations like the ‘Heroic Failure Award’ from Proctor & Gamble (Morgan 2015) while others have positive connotations like the ‘Dare to Try’ award by Tata (Waczek 2012). Nonetheless, the concept remains unchanged. For simplicity, we keep referring to the award as a ‘Failure Award’.

[8] On average, participants receive a total compensation of $4.36 for completing the study in approx. 31 minutes. The compensation was above the average MTurk reservation wage of $1.38 per hour (Horton & Chilton, 2010).

[9] In case participants decide on project continuation, a computer randomly selects whether the best-case scenario or the worst-case scenario of the project’s outcome occurs.

provided with a project account for the initial investment of 5 m lira. Participants' variable compensation is 1% of the balance of the project account at the end of the experiment.[10]

Participants in the Failure Award treatments are then introduced to the Failure Award manipulation that we describe below. After completing a comprehension quiz, participants start the (3) risk task. Here, participants have to select one of two innovative projects, which differ in terms of risk and return. After selecting a project, participants enter the (4) EoC task.[11] They learn that one year has passed and receive negative updates regarding the selected project. Participants must decide whether to continue investing in their initial (but failing) project or to invest in a safe alternative project with a certain and higher expected payoff. Participants who continue the initial project receive another negative project update and must provide a final decision about the continuation of the project.[12] Finally, participants respond to a (5) post-experiment questionnaire and are informed about their compensation.

3.2 Manipulation

Participants in the Failure Award treatments learn that the management board started to grant Failure Awards. An example of a recent Failure Award winner is presented, and a video is embedded that illustrates an excerpt from an award ceremony. The two alternative Failure Award type manipulations are presented in Figure 2. Participants in the innovation type treatment are told that managers often shy away from "taking risks and being innovative" when facing difficult decisions. Therefore, CleverClean has started granting Failure Awards to managers who do not shy away but have the courage to "take the risk to start a highly innovative

[10] A project account balance is included to link the participant's compensation to the project's success. Due to the variable compensation of 1% of the projects account balance, 1 m lira can be converted to a payout of 0.50 dollar.
[11] For our EoC analyses, only participants selecting the riskier project are relevant, since granting the Failure Award is dependent on this choice. However, due to fairness reasons we also let participants selecting the safer project finish the experiment and paid them accordingly.
[12] A second EoC round is implemented to create a more realistic scenario in which participants can delay their project termination decision and justify it by relying on the opportunity to end it at a later point if it turns out that the project keeps failing.

project". In the discontinuation type treatment, participants are told that managers shy away from "'pulling the plug' of a failing project". Thus, Failure Awards are granted to managers who do not shy away but "'pull the plug' and stop wasting resources by terminating a failing project". We modeled our Failure Award types after practical examples (e.g., Google X) (Leber 2016).[13]

[Insert Figure 2 about here]

The criteria to receive a Failure Award are kept constant across the two types of conditions. Participants receive a Failure Award if they start a risky project and deliberately terminate the project as soon as failure becomes imminent. Practical examples show that Failure Awards have a symbolic meaning and are often nonmonetary (e.g., trophies, applause, award ceremonies) or have a symbolic cash component (e.g., Google X) (Johnson 2017; Stewart 2015). Hence, participants in the Failure Awards conditions learn that in addition to a reward ceremony, award winners receive 2,000 lira, which equals \$0.10 (approx. 2% of the average total compensation).[14] To rule out that the (minuscule) compensation associated with the Failure Award drives our effects, we hold the difference in the expected compensation between the failing project and the alternative (economically preferable) investment option constant across all conditions. More precisely, since participants in the Failure Award absent condition do not receive a Failure Award, the expected return of the alternative project is 0.2 m lira greater (equaling a compensation of 2,000 lira) than the expected return of the Failure Award

[13] To differentiate the provided Failure Award type manipulation from a goal-setting manipulation (Kachelmeier et al. 2016), all treatments receive the information that the companies' goals are to engage in innovations through risk-taking and to reduce resource wastage in failing projects. Consequently, the Failure Award types serve as a supplementary control mechanism that provides a cue suggesting appropriate behavior by additionally rewarding this behavior (Kachelmeier et al. 2016).

[14] Participants are told that the company has a Failure Award budget of 0.2 million lira of which every participant who qualifies will receive 1%.

treatments. In effect, all treatments have the same expected compensation from project continuation (4.32 m) or discontinuation (5.0 m vs. 4.8 + 0.2 m).[15]

3.3 Task Description

Risk Task

The risk task is modeled after the choice problems from Kahneman and Tversky (1979). Individuals choose between option A (B), with a high (lower) variance and a high (lower) expected outcome (Laux et al. 2018).[16] In the experiment, the safer project (B) introduces the Smart Vacuum Robot, an artificially intelligent vacuum robot (Figure 3). Based on currently known information, predicted values indicate that with a 90% (10%) probability, the project results in expected discounted cash inflows of 7 m (2 m) (million lira).[17] This leads to an expected return of 6.5 m with a variance of 2.25.[18] To summarize this financial information, an investment rating of two out of five stars is presented. It is emphasized that this project does not qualify for a Failure Award.[19] The riskier project (A) is represented by the Smart Mop Robot, an artificially intelligent vacuum and mop robot (Figure 4). The project leads to predicted cash inflows of either 14 m or 0 m lira with equal probabilities. This equals an expected return of 7 m (compared to 6.5 m of the safer project) and a variance of 49 m (compared to 2.25 m). The investment rating is four out of five stars. Participants are informed that this project qualifies for a Failure Award.

[Insert Figure 3 and 4 about here]

---

[15] Participants indicate an average value of 2.8 ($p < 0.01$ when tested against the scale mid-point) to the question "The monetary compensation of $0.10 (2,000 lira) from the Failure Award was important to me."; 1 = not important, 7 = highly important), which proves that the Failure Award is effective through its intrinsic value.

[16] In the study from Kahneman and Tversky (1979) a 100% certain option is provided. However, to subsequently draw implications for firms a more realistic scenario is chosen in which the outcome variance is minimal but not equal to zero.

[17] Here and in the following, all values in million units are reported in lira unless stated otherwise.

[18] For both projects, it is made clear to the participant that the values are predicted based on currently known information and that project cash inflows are expected. Thus, all values are subject to change.

[19] As shown in practice, Failure Awards are not granted for every project that fails, but rather for projects that are truly innovative and require a substantial amount of risk (e.g., Proulx 2019). Thus, a project with a predicted certainty of 90% is unlikely to be eligible for a Failure Award.

Participants are told that the company prefers projects with higher expected returns. Comparing the two projects, the riskier project is economically preferred due to its higher expected value.[20] Risk-taking – the first dependent variable – is measured through the binary choice of which project participants select.[21]

Escalation of Commitment Task

To induce an EoC setting, decision-makers have to receive negative feedback on their initial decision (Wong et al. 2006). Seybert (2010), Brink et al. (2020) and Denison (2009) provide negative information to their participants after the initial decision, which shows a decline in expected cash flows. Similarly, our participants receive negative project feedback after making the first investment decision (i.e., project choice). Hence, participants are informed that the development process is worse than initially expected (e.g., lower expected sales due to a new competitor) and that additional investments are required.[22] Updated predicted financials indicate that the project's expected return decreased from 7 m to 4.32 m. Participants are also informed that they can invest the remaining funds in an alternative project that yields higher and certain expected returns (Brink et al. 2020; Seybert 2010). More precisely, the alternative project generates expected cash flows of 5 m in the Failure Award absent and 4.8 m in the Failure Award present condition. Additionally, participants in the Failure Award present condition receive 0.2 m from the Failure Award. Consequently, all participants have the same basis for their variable compensation (5.0 m vs. 4.8 + 0.2 m). For all treatments, the alternative project generates higher expected returns (4.8 and 5.0 m) than continuing the initial project (4.32

[20] This is in line with expected utility theory, which states that (risk-neutral) individuals should make decisions based on expected returns and therefore always choose the option with higher expected returns independent from the inherent risk (i.e., variance) Kahneman and Tversky (1979); Schoemaker (1982).
[21] Besides measuring risk-taking through the project choice, another purpose is to establish a personal responsibility for the project, which facilitates escalation of commitment (Denison 2009; Schoorman and Holahan 1996).
[22] The following description refers to the riskier project since the Failure Award depends on the selection of the riskier project. Hence, the effect of Failure Awards on EoC can only be measured through participants who chose the riskier project.

m). Thus, the unambiguous 'correct' answer is to invest in the alternative project and abandon the initial project. Participants are informed that if they decide to continue their project, they will be able to terminate it after another year has passed.[23] However, a delayed termination of the failing project reduces the chance of receiving a Failure Award to 50%.[24]

Finally, participants are asked to indicate to the management board their willingness to continue the project on a 101-point scale (0% = project termination, 100% = project continuation) (Keil et al. 2000; Wong 2005). A rational decision-maker would indicate a value of 0% and thereby recommend terminating the project (Harrell and Harrison 1994). Correspondingly, EoC – our second dependent variable – is measured by the percentage of project continuation. Impression management concerns and self-justification pressure are deliberately provoked in this study to measure EoC. Therefore, participants are told that their decisions will be reviewed by the experimental administrator, and in some cases, they might receive written feedback on their decision-making process. In this vein, approx. 5% of the participants were randomly selected and received a message through MTurk that evaluates the rationality of their decisions.

Directly after the recommendation, participants are asked to make a binary decision whether to continue or terminate the project (Brink et al. 2020).[25] The Failure Award is granted based on this decision. In case participants decide to terminate their project, the EoC task ends. Otherwise, they enter a second EoC round. The second round is similar to the first round, with the only differences being that the project outcome probabilities worsen (Behrens and Ernst 2014) and that the award is granted with only a 50% probability in case of project termination.

[23] Brockner (1992) argues that "[…] escalation situations include repeated (rather than one-shot) decision-making in the face of negative feedback […]", Brockner (1992), p. 40. Since we expect that Failure Awards might lead to a delayed termination, an additional EoC round with a further decreasing expected value is incorporated. We approximate EoC by using participants' first continuation decision, since project termination is the economically preferred decision at this point of time. Nevertheless, we run additional analyses based on participants' delayed project continuation decisions.

[24] This reduced likelihood of receiving a Failure Award is implemented, since a delayed project termination contradicts the objective of a Failure Award, which aims at saving resources by terminating failing projects as soon more appropriate resource usage seems feasible.

[25] Using the binary variable to measure EoC does not change the general significance of our main results.

3.4 Participants

Participants were recruited from the MTurk internet marketplace through a publicly advertised Human Intelligence Task (HIT). MTurk offers an easily accessible and cost-efficient platform (Brasel et al. 2016; Paolacci et al. 2010) that provides reliable data, especially due to its diverse participant pool (Buhrmester et al. 2011; Hunt and Scheetz 2019). Moreover, MTurk workers are more representative of the U.S. population in terms of demographics, behavioral patterns and risk preference attributes than undergraduate students (Buhrmester et al. 2011; Farrell et al. 2017; Goodman et al. 2013). This allows for greater generalizability of the study's results. Furthermore, MTurk workers demonstrate a similar susceptibility to cognitive biases to that of participants in laboratory experiments.

Based on Bentley's (2021) four sources of noise in MTurk research, we took precautionary steps by prescreening the population. Hence, workers were eligible to participate in the study only if they had a historical HIT approval rating of 95% or higher, completed at least 500 HITs and were based in the U.S. (Peer et al. 2014). Several questions, including two attention check questions based on Peer et al. (2017) and Liu et al. (2020), were included to ensure that participants understood the experiment and were attentive throughout the PEQ.[26] Furthermore, using mobile devices was prohibited to minimize possible distractions. Last, if participants spent less than the minimum required time on a page based on minimal page times collected during the pretest, they could not proceed with the experiment (Hunt and Scheetz 2019).

In total, 264 participants successfully completed the experiment. Participants' average age was 40.3 years, 37.1% were female, and approx. 84% had a bachelor's degree or higher. Furthermore, 215 (81%) participants had six years or more of work experience. Based on the ex-ante risk-elicitation task, we found that 63.6% of the participants were risk-averse, 16.7%

[26] From the 277 subjects who finished the experiment, 13 failed at least one of the two attention check items and are excluded from the following analyses.

were risk-neutral, and 19.7% were risk-seeking. This is in line with previous research that finds a preference for risk aversion among individuals (Crosetto and Filippin 2013; Kreilkamp et al. 2021). Finally, there are no significant differences across conditions for age, gender, risk preferences, working experience, educational degree, prior knowledge of biases or of Failure Awards (all p-values > 0.21).[27] Hence, randomization was successful.

4 Results

4.1 Comprehension Checks

Before testing our hypotheses, we check participants' understanding of the task. To create a valid EoC setting, participants need to comprehend that their initially chosen project is failing. Thus, we ask participants in the post-experimental questionnaire on a 7-point Likert scale to what extent they agree with the following statement: "According to CleverClean, continuing the project meant to invest more money in a failing project" (1 = totally disagree, 7 = totally agree). In all treatments, the mean is above the scale midpoint ($p < 0.016$). Hence, subjects understood that their project was failing. Furthermore, we check whether Failure Awards create a culture in which participants perceive that failure is tolerated (i.e., open EMC). We find that subjects in the Failure Award treatments agree more with the statement "I feel that at CleverClean, failures are tolerated and not punished" than those in the Failure Award absent treatment ($t = -8.17$, $p < 0.01$). Hence, Failure Awards can be used to create an open EMC.

Moreover, we check whether participants correctly identified the risker project. Therefore, participants indicated which project they believe to be riskier (1 = Smart Vacuum Robot, 7 = Smart Mop Robot). With a mean of at least 6.36, participants correctly identified the Smart Mop Robot as the riskier project in all treatments.[28] Last, participants in both Failure Award conditions correctly selected the three conditions required to qualify for a Failure Award (i.e.,

[27] All p-values are reported as two-tailed unless stated otherwise.
[28] Excluding the 11 subjects who indicated a value of 4 or less leads to inferentially identical results.

starting a risky project, project failure and project termination). Hence, participants in both treatment groups knew equally well the criteria to receive a Failure Award.

4.2 Descriptive Results and Hypotheses Tests

Table 1, Panel A shows and Figure 5 Panel A illustrates the descriptive statistics for the project choice of the dependent variable risk-taking. Consistent with H1a, risk-taking is higher when discontinuation-type Failure Awards are granted (74%) than when Failure Awards are absent (47%). The difference in risk-taking between the discontinuation (74%) and innovation (73%) types is rather small.

[Insert Table 1 and Figure 5 about here]

An analysis of variance (ANOVA) shows that risk-taking significantly differs between the three treatments (Table 2, Panel A, $F = 10.18$, $p < 0.01$). We use planned contrasts (Guggenmos et al. 2018; Rosenthal et al. 2011) to test whether risk-taking differs between both Failure Award treatments and the Failure Award absent treatment. Therefore, we assign contrast weights of +1 to the discontinuation and innovation type treatments to test their cell mean against the Failure Award absent condition, which is assigned a contrast weight of -2 (Table 2, Panel B). The contrast is positive and significant (contrast = 0.53, $p < 0.01$), indicating that Failure Awards significantly increase risk-taking compared to the Failure Award absent condition.[29] For the formal test of H1a, we apply pairwise comparisons. The results in Table 2, Panel C show that participants in the discontinuation type take significantly more risk than participants in the Failure Award absent condition ($t = 3.83$, $p < 0.01$).[30] Hence, H1a is supported.

[29] Testing for sample inequality (i.e., cell C is lower than the average of cells A and B) does not require a test of the residual between-cells variance (Guggenmos et al. 2018).
[30] Using the discontinuation type as the baseline in a linear OLS regression model (untabulated), the discontinuation treatment also significantly differs from the Failure Award absent treatment ($t = 3.83$, $p < 0.01$) and does not differ from the innovation treatment ($t = -0.11$, $p = 0.910$). To estimate treatment effects on binary outcomes, applying linear OLS regression models is generally more appropriate than using logit models (Gomila 2021). Nevertheless, logit regression coefficients confirm these results ($z = 3.53$, $p < 0.01$; $z = -0.12$, $p = 0.905$).

In line with H1b, the descriptive results show that risk-taking is higher when innovation-type Failure Awards are granted (73%) than when Failure Awards are not granted (47%). The pairwise comparisons in Table 2, Panel C show that risk-taking is significantly higher in the innovation-type Failure Award condition ($t = 3.78$, $p < 0.01$). Thus, H1b is supported. Lastly, we find no difference between discontinuation-type Failure Awards and innovation-type Failure Awards ($t = 0.11$, $p = 0.91$). We revert to this finding in the additional analyses section.

[Insert Table 2 about here]

Looking at EoC[31], descriptive results in Table 1, Panel B show that the likelihood of project continuation is lower when discontinuation-type Failure Awards are granted (43.96%) than when Failure Awards are absent (62.68%). This is consistent with H2. We again applied an ANOVA, which shows that EoC significantly differs across all three treatment groups (Table 3, Panel A, $F = 3.16$, $p = 0.045$). To formally test H2, we use pairwise comparisons (Table 3, Panel B). Our results show that EoC is significantly lower in the discontinuation type compared to the Failure Award absent treatment ($t = -2.51$, $p = 0.013$). Hence, H2 is supported.

The research question RQ2 examines whether EoC is lower in the innovation-type Failure Award treatment (53.24%) versus the Failure Award absent condition (62.68%). Even though descriptive results suppose a reducing effect, pairwise comparisons reveal that the innovation type treatment does not significantly differ from the Failure Award absent treatment (Table 3, Panel B, $t = -1.28$, $p = 0.201$). We further analyze this finding in the additional analyses section. Figure 5, Panel B illustrates the results.

[Insert Table 3 about here]

[31] As only participants who selected the riskier project were able to receive a Failure Award, the sample reduces from 264 to 165 participants for the EoC analyses.

4.3 Additional Analyses

Factor Analysis of Psychological Safety

This section presents additional analyses by leveraging items from the post-experimental questionnaire (PEQ) to substantiate our theory. In our hypothesis, we predict that Failure Awards increase decision-makers' level of psychological safety. More precisely, we argue that psychological safety has a twofold effect. First, we predict that Failure Awards –independent of their type– increases the feeling of being safe to admit failures (PS-I). Second, we argue that the innovation-type award increases the feeling of being safe to experiment and take risks (PS-II). Therefore, we apply principal component analysis to extract factors based on psychological safety items, which cover participants' perceptions of feeling safe to admit failures (PS-I, Table 4, Panel A) and of feeling safe to take risks and experiment (PS-II, Table 4, Panel B).

[Insert Table 4 about here]

The PS-I factor (PS-II factor) has an eigenvalue of 2.11 (1.88) and a Kaiser–Meyer–Olkin (KMO) of 0.62 (0.502).[32] Using pairwise comparisons (Table 5, Panel A) we find that participants feel significantly safer to admit failure (PS-I) when Failure Awards are present compared to when they are absent ($t = 6.26$, $p < 0.01$). We find this positive effect for both Failure Award types (innovation type: $t = -6.03$, $p < 0.01$ and discontinuation type: $t = -4.51$, $p < 0.01$).[33] As argued, we do not find a significant difference in PS-I between the two Failure Award types ($t = -1.33$, $p = 0.186$). These findings support our theory that Failure Awards – irrespective of the framing type– induce psychological safety through which individuals feel safe to admit failure.

Looking at the PS-II factor (Table 5, Panel B), we find that participants in the innovation type condition feel significantly safer taking risks compared to participants in the Failure

[32] A minimum KMO of 0.5 is necessary for reliable factor analytic estimation (Kaiser 1970).

[33] These and the following results of the PEQ items are based on the full sample ($n = 264$) since all participants are exposed to the Failure Award manipulation before selecting their project.

Award absent condition ($t = -2.06$, $p = 0.041$) and in the discontinuation type condition ($t = -1.71$, $p = 0.089$). Yet, we do not find that participants in the discontinuation type treatment feel significantly safer taking risks compared to the Failure Award absent group ($t = -0.20$, $p = 0.839$). Hence, in line with our prediction, the second effect of psychological safety –feeling safe to take risks and experiment– is only triggered in the innovation type treatment. This explains why we find an EoC reducing effect for the discontinuation but not for the innovation type.

Although the innovation type explicitly triggers the feeling of being safe to take risks, we do not find a significant difference between the two Failure Award types on risk-taking ($t = 0.11$, $p = 0.91$; Table 2, Panel C). To explain this finding, we extract a third factor which, in contrast to the PS-II factor, is based on items measuring risk perception (i.e., participants' perception of feeling safe to take risks) immediately after participants made their initial project decision.[34] This allows us to examine how participants perceived their environment at the time risk-taking was measured. This factor has an eigenvalue of 1.81 and an overall KMO of 0.57 (Table 5). Contrary to the PS-II factor, we do not find a significant difference in risk perception between the two Failure Award types ($t = -0.05$, $p = 0.964$, untabulated). Thus, at the beginning of the experiment, the discontinuation type has a similar effect on participants' risk perception compared to the innovation type.

We explain the differences in participants' risk perception by their emotional attachment to the project (Druckman and McDermott 2008; Giorgi 2017). At the time participants select their project (i.e., when we measure risk-taking), they are less emotionally attached to it. Due to the emotional distance, the specific type of the Failure Award is less effective (Druckman and McDermott 2008), which is why participant's perceived risk does not differ between

[34] In contrast, the PS-II factor was measured after the EoC task in the PEQ at the end of the experiment.

the two types. Consequently, participants are more likely to objectively evaluate their decision and select the economically preferred project (i.e., the risky project). Once the experiment continues, participants become more invested in the project, which increases their emotional attachment (Hennig et al. 2023). This in turn increases the effectiveness of the specific type of Failure Award. That is why we find that participants in the innovation type treatment feel safer taking risks when we measure risk perception at the end of the experiment (PS-II).

Overall, we find consistent with our predictions that both psychological safety factors are positively linked to risk taking (PS-I: $t = 3.80$, $p < 0.01$; PS-II Risk: $t = 7.99$, $p < 0.01$, untabulated). On the contrary, the feeling of being safe to admit failures (PS-I) is negatively ($t = -2.36$, $p = 0.019$, untabulated) related to EoC, whereas the feeling of being safe to take risks (PS-II EoC) is positively linked to EoC ($t = 10.65$, $p < 0.01$, untabulated).

[Insert Table 5 and 6 about here]

Sunk Costs and Escalation of Commitment

Brockner et al. (1981) show that sunk costs significantly influence escalation behavior. We use the following item from Brockner et al. (1981) to measure the relevance of sunk costs: *"I had already invested so much that it seemed silly… 1 = to spend another penny to 7 = not to invest a little more"*. Higher values indicate an increased sensitivity to the sunk cost effect.

The mean value for both Failure Award treatments combined is 3.33 and 4.09 for the Failure Award absent treatment, with a significant difference between the two treatments ($t = - 3.35$, $p < 0.01$). This indicates that Failure Awards significantly decrease participants' sensitivity to sunk costs. Further, we find that the sensitivity to sunk costs significantly increases the likelihood of project continuation ($t = 9.56$, $p < 0.01$). The presence of Failure Awards decreases participants' sensitivity to sunk costs since timely project termination is supported by the organization, decreasing the perceived need to explain why organizational resources were wasted (Arkes and Blumer 1985; Sleesman et al. 2012). Thus, Failure Awards indirectly decrease EoC by reducing decision-makers' sensitivity to sunk costs.

The Impact of Failure Awards on Delayed Project Termination

Failure Awards induce the feeling of being safe to take risks and experiment, which could lead to a delayed termination decision instead of immediate termination. Thus, our experimental design incorporates a second decision round in which participants receive a second project update indicating a slightly lower expected return after they decide to continue their already poorly performing project. Participants knew that a Failure Award is granted with only a 50% probability in case of delayed project termination.

ANOVA results show no difference between the three treatments concerning delayed EoC ($F = 0.540$, $p = 0.727$, untabulated). Using pairwise comparisons, none of the three comparisons shows a significant difference between the treatment pairs (all p-values > 0.43, untabulated). Hence, discontinuation-type Failure Awards immediately decrease escalation tendencies (H2), but they do not decrease discontinuation tendencies in case decision-makers delay their discontinuation decision ($t = -0.28$, $p = 0.779$, untabulated).

5 Conclusion

The "fear of failure" harms firms' position in today's competitive environment. On the one hand, it increases employees' reluctance to take risks when making investment decisions. On the other hand, the fear of failure leads to employees falling prey to Escalation of Commitment (EoC), i.e., the tendency to overinvest in failing projects (Staw 1976). To counteract these issues, a growing number of organizations grant Failure Awards to employees who started risky but economically preferred projects that ultimately failed. Usually, firms employ (1) an innovation type, which focuses on encouraging risk-taking, or (2) a discontinuation type, which focuses on terminating failing projects (i.e., reducing EoC) when creating this award. Our study tests how Failure Awards and their different types affect EoC and risk-taking.

We conducted an online experiment using MTurk in which participants are first asked to invest either in a safer or riskier project (our proxy for risk-taking). Next, they decide whether to terminate their project when learning it is failing (our proxy for EoC). Only if participants choose the riskier project and terminate it, they can receive the highest expected compensation, including a Failure Award.

We provide evidence that the presence of Failure Awards and their type affect risk-taking and EoC. Specifically, we find that risk-taking is significantly encouraged through both types of the Failure Award. Further, we find that discontinuation-type Failure Awards decrease EoC. However, we do not find this de-escalating effect for innovation-type Failure Awards. We attribute our findings to the Behavioral Agency Model (Wiseman and Gomez-Mejia 1998) and identify psychological safety (Edmondson 1999) as an underlying factor.

We predict and find that Failure Awards induce psychological safety, which mitigates the fear of failure. However, psychological safety has a twofold effect on employees. On the one hand, it creates a feeling of being safe to admit failures, and on the other hand, it creates a feeling of being safe to experiment and take risks. The latter effect is particularly triggered through the innovation type. Whereas both effects of psychological safety enhance risk-taking, the influences on EoC are opposing. Thus, through the feeling of being safe to experiment and take risks, individuals with an innovation-type Failure Award tend to take the risk to further invest in their failing project, which reduces the general de-escalating effect of Failure Awards.

Our findings have important implications for the design of management control systems. First, our findings illustrate that Failure Awards encourage risk-taking, independent of their type. Second, we show that discontinuation-type Failure Awards can be used as a cost-efficient alternative to reduce EoC. Our results imply that it is crucial to pay close attention to what aspects Failure Awards highlight, as a focus on innovation and risk-taking, which can be predominantly found in practice (e.g., "Epic Fail Award" by Hill Holiday (Proulx 2019), "Heroic Failure Award" by Proctor & Gamble (Anthony 2020) or "Lean Forward; Fail Smart Award" by NASA (NASA 2021) does not reduce EoC. Third, we provide evidence that psychological safety is a driving factor between Failure Awards and decision-makers' risk-taking and escalation behavior. Referring to Barton et al. (1989), who do not find a decrease in EoC through employing an open error management climate, discontinuation-type Failure Awards seem to overcome this challenge by actively incentivizing project discontinuation.

Future research might further explore this field of research. While we implement a Failure Award with a small monetary value, future research should investigate whether the results hold when completely non-monetary Failure Awards are employed. Even though we can rule out that the effectiveness of Failure Awards is driven by the monetary component, as our design keeps the relative payout constant between all treatments, it would be interesting to see whether granting a trophy or applause has stronger effects within a non-online, e.g., a laboratory, setting. Additionally, whereas in our setting early abandonment of the failing project is rational and thus preferred, one could investigate whether Failure Awards can also lead to irrational early abandonment of well-performing projects.

Moreover, our study focuses on the two extremes of Failure Award types which either highlight innovation or discontinuation. Future studies could examine the effects of a Failure Award type that highlights both aspects equally. Finally, it could be examined whether Failure Awards as a potential de-escalation tool are transferable to other biases, e.g., overconfidence (Malmendier and Tate 2005). Since Failure Awards turn mistakes into "something less negative", an individual's overly optimistic self-assessment, also known as *overconfidence bias* (Moore and Healy 2008), might be attenuated.

Table 1

Descriptive statistics (mean, [standard deviation])

	Failure Award Type[a]				
	Innovation type	Discontinuation type	***Total***	Failure Award absent	**Total**
Panel A: Risk-taking behavior (n = 264)					
Number of subjects	81	76	**157**	107	**264**
Choice of risky project[b]	0.73 [0.45]	0.74 [0.44]	**0.73 [0.44]**	0.47 [0.50]	**0.63 [0.49]**
Panel B: Escalation of Commitment (n = 165)					
Number of subjects	59	56	**115**	50	**165**
Willingness of project continuation[c] – risky project	53.24 [38.73]	43.96 [38.99]	**48.72 [38.97]**	62.68 [36.91]	**52.95 [38.78]**

[a] The *type of the Failure Award* is manipulated between subjects at two levels. In the innovation type treatment, participants are told that Failure Awards are granted to managers who have the courage to "take the risk to start a highly innovative project". In contrast, in the discontinuation type treatment, participants are told that Failure Awards are granted to managers who have the courage to "'pull the plug' and stop wasting resources by terminating a failing project."

[b] *Choice of risky project* [0: safe project, 1: risky project] represents the number of participants who chose the risky project *Smart Mop Robot* with an expected value of 7 m and a variance of 49 (SD of 7.0) compared to the safe project *Smart Vacuum Robot* with an expected value of 6.5 m and a variance of 2.25 (SD of 1.5).

[c] *Willingness of project continuation* represents the indicated percentage (on a 101-point scale with 0% = termination and 100% = continuation) to which participants are willing to continue their initially chosen but poorly performing project. As we measure the effect of Failure Awards on EoC a prerequisite is that participants have the chance to receive a Failure Award. This is only the case if participants chose to initiate the riskier project, which is why this table reports the results only for participants who chose the riskier project (n = 165).

Table 2

Effects of Failure Awards on Risk-Taking[a]

Dependent variable: Choice of risky project (n = 264)

Panel A: ANOVA Model

Source of variation	df	MS	F-statistic	p-value[c]
Treatments[b]	2	4.48	10.18	<0.01
Error	261	57.40		
Total	263	61.88		

Panel B: Planned Contrasts

Source of variation	F-statistic	p-value
Discontinuation type/Innovation type/ No Failure Award [+1, +1, -2]	20.36	<0.01

Panel C: Pairwise Comparisons

Treatments	t-statistic	p-value
Discontinuation type > No Failure Award [H1a]	3.83	<0.01
Innovation type > No Failure Award [H1b]	3.78	<0.01
Discontinuation type < Innovation type	0.11	0.91

[a] The dependent variable risk-taking is operationalized through the *choice of risky project*, a binary variable with 0 = choice of the safer project and 1 = choice of the riskier project. The riskier project is the *Smart Mop Robot* project, with an expected value of 7 m and a variance of 49 (SD of 7.0). The safe project is the *Smart Vacuum Robot* project with an expected value of 6.5 m and a variance of 2.25 (SD of 1.5).

[b] The variable *Treatments* is separated into the three groups: 1) Discontinuation type, 2) Innovation type and 3) No Failure Award.

[c] All p-values are reported as two-tailed.

Table 3

Effects of Failure Awards on Escalation of Commitment[a]

Dependent variable: Willingness of project continuation – risky project[b] (n = 165)

Panel A: ANOVA Model

Source of variation	df	MS	F-statistic	p-value[d]
Treatments[c]	2	4630.06	3.16	0.045
Error	162	1465.39		
Total	164	1503.99		

Panel B: Pairwise Comparisons

Treatments	t-statistic	p-value
Discontinuation type < No Failure Award [H2]	-2.51	0.013
Innovation type < No Failure Award [RQ2]	-1.28	0.201
Discontinuation type < Innovation type	-1.30	0.196

[a] The dependent variable Escalation of Commitment is operationalized through the *willingness of project continuation*, which represents the indicated percentage (on a 101-scale with 0% = termination and 100% = continuation) to which participants are willing to continue their initially chosen but poorly performing project.
[b] Since only participants who chose the riskier project are eligible to receive a Failure Award, the sample reduces to 165 for the EoC measurement. Due to fairness reasons, all other participants were still able to finish the experiment and receive the compensation.
[c] The variable *Treatments* is separated into the three groups: 1) Discontinuation type, 2) Innovation type and 3) No Failure Award.
[d] All p-values are reported as two-tailed.

Table 4

Factor Analyses of the Construct of Psychological Safety

Panel A: "Safety to admit failure" Items (PS-I)

<u>Questions (7-point scale)</u>

1. I feel that at CleverClean, failures are tolerated and not punished.
(endpoints: totally disagree and totally agree)

2. I feel that at CleverClean, mistakes are perceived as an opportunity to improve oneself.
(endpoints: totally disagree and totally agree)

3. To what extent do you feel the need to justify your initial project decision? [a]
(endpoints: not at all and very strong)

4. In your opinion, what is the likelihood that terminating the project results in negative personal consequences (e.g., decreased promotion probability): [a]
(endpoints: not likely at all and very likely)

5. I was afraid that important persons (e.g., superiors) could receive a bad impression of me in case I terminate the project. [a]
(endpoints: totally disagree and totally agree)

6. I thought that it would make a good impression if I…" [a]
(endpoints: terminate the project and continue the project)

7. I am afraid to receive negative feedback from the experimental administrator. [a]
(endpoints: totally disagree and totally agree)

Note: The questions are based on Edmondson (1999), Roetzel et al. (2020), Brink et al. (2020), Steinkühler et al. (2014) and Brockner et al. (1981).

Panel B: "Safety to take risks and experiment" Items (PS-II)

<u>Questions (7-point scale)</u>

1. In my role as a manager at CleverClean I had concerns about taking risks. [a]
(endpoints: totally disagree and totally agree)

2. How would you characterize the decision to continue the project?
(endpoints: significant threat and significant opportunity)

3. How would you characterize the decision to continue the project?
(endpoints: potential for loss and potential for gain)

4. I feel that at CleverClean, mistakes are perceived as an opportunity to improve oneself.
(endpoints: totally disagree and totally agree)

Note: The questions are based on Edmondson (1999), Sitkin and Weingart (1995) and Wong (2005)
[a] Marked items have been reversed for computing the factor.

Table 5

Factor Analysis on Psychological Safety

PS-I - "Safety to admit failure"

Panel A: Pairwise Comparisons

Treatments	t-statistic	p-value
No Failure Award < Failure Award (both types)	6.26	<0.01
No Failure Award < Innovation type	-6.03	<0.01
No Failure Award < Discontinuation type	-4.51	<0.01
Discontinuation type < Innovation type	-1.33	0.186

PS-II - "Safety to take risks and experiment"

Panel B: Pairwise Comparisons

Treatments	t-statistic	p-value
No Failure Award < Failure Award (both types)	-1.36	0.175
No Failure Award < Innovation type	-2.06	0.041
No Failure Award < Discontinuation type	-0.20	0.839
Discontinuation type < Innovation type	-1.71	0.089

Note: All p-values are reported as two-tailed and n = 264.

Table 6

Factor Analysis of the Construct of Risk Perception

Risk perception was measured right after participants chose their project

Questions (7-point scale)

1. How would you characterize your selected project?
(endpoints: negative situation and positive situation)

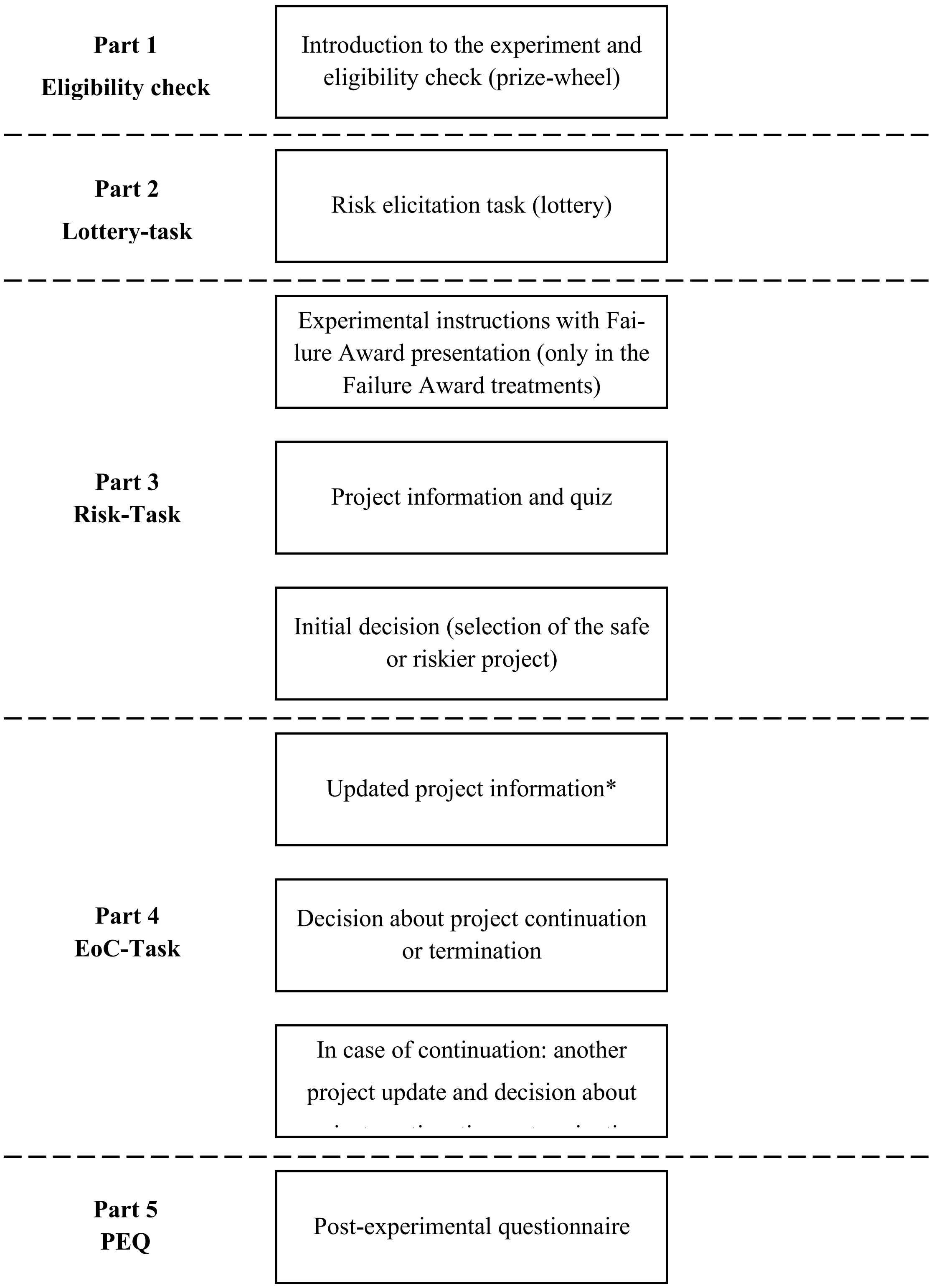
2. How would you characterize your selected project?
(endpoints: potential for loss and potential for gain)

3. What is the likelihood of your chosen project to succeed?
(endpoints: very unlikely to very likely)

Note: The questions are based on Sitkin and Weingart (1995) and Wong (2005)

Figure 1

Experimental Procedure



* Due to fairness considerations, we also let participants choosing the safe project finish the experiment and compensated them accordingly.

Figure 2

Failure Award Type Manipulation
(differences are printed in bold)

Innovation Type	Discontinuation Type
CleverClean is one of the first companies that implemented a new type of reward for its managers - the ***Courage Award***.	
What does courage mean for CleverClean? *CleverClean* understands that courage is required to run a successful business. Its managers face difficult decisions every day, - and it often takes a lot of courage to make the 'right' decision. For example, managers often shy away from **taking risks and being innovative**. This is where the Courage Award comes into play. *CleverClean* now awards managers who do not shy away but **take the risk to start a highly innovative project**. Obviously, the management knows that even good ideas may fail. Thus, in case you **do not shy away but start a project which implies a substantial amount of risk and appears innovative, *CleverClean* supports you with the Courage Award.** Of course, you do not receive this award for every risky project you start. You only receive this supporting award if the risky project is failing and you decide to discontinue it.	***What does courage mean for CleverClean?*** *CleverClean* understands that courage is required to run a successful business. Its managers face difficult decisions every day - and it often takes a lot of courage to make the 'right' decision. For example, managers often shy away from **'pulling the plug' of a failing project**. This is where the Courage Award comes into play. *CleverClean* now awards managers who do not shy away but **'pull the plug' and stop wasting resources by terminating a failing project**. Obviously, the management knows that even good ideas may fail. Thus, in case you **do not shy away but 'pull the plug' of a project to save resources, *CleverClean* supports you with the Courage Award**. Of course, you do not receive this award for every project you discontinue. You receive this supporting award only if the discontinued project is failing and it implied a substantial amount of risk and appeared innovative when started.
Taylor is the most recent winner of the *Courage Award*. Take a look at Taylor's achievement:	
Taylor received the *Courage Award* for **taking the risk** to start an innovative project which focused on developing a cleaning product for universal usage. Unfortunately, it turned out that the overall product won't be profitable. *CleverClean* supported Taylor's courage of **taking the risk** to start the project by granting the Courage Award, after Taylor terminated the failing project.	Taylor received the *Courage Award* for starting an innovative project which focused on developing a cleaning product for universal usage. Unfortunately, it turned out that the overall product won't be profitable. *CleverClean* supported Taylor's courage to **'pull the plug'** of the failing project by granting the Courage Award for the termination of the project.
The following clip shows the latest award ceremony, where a manager received a Courage Award for showing the courage to **take risks**:	The following clip shows the latest award ceremony, where a manager received a Courage Award for showing the courage to **'pull the plug'**:

Figure 3

Introduction of the Smart Vacuum Robot (safe project)

Project idea #1 | Project idea #2

Project idea #1
Smart Vacuum Robot

Clean clever

Clean

Clever

with CleverClean Inc.

Project overview

Description

Risk: ●●○○○
Return: ●●○○○
Qualifies for Courage Award: **No**

Smart Vacuum Robot is an artificially intelligent vacuum robot. It has a noise cancelling function which makes the robot so quiet that even babies can keep sleeping while the house gets vacuumed.

Its new feature is a **mapping function** through which the robot can be sent to any place via an app.

CleverClean has successfully launched a similar vacuum robot last year, but without the **mapping and noise cancelling functions**. Based on the profound experience with the previous vacuum product and the available market surveys, the returns for Smart Vacuum Robot will be **modest but very certain**.

The initial investment to start the Smart Vacuum project amounts to **3m Lira**.

Financials

The current balance of the project account is 5m Lira.

The following **predicted values** were calculated based on currently known information:

	Best-Case Scenario	**Worst-Case Scenario**
Probability	**90%**	**10%**
Required initial investment	3m Lira	3m Lira
Expected project cash inflows	7m Lira	2m Lira

Figure 4

Introduction of the Smart Mop Robot (risky project)

Project idea #1 | Project idea #2

Project idea #2
Smart Mop Robot

Project overview

Description

Risk: ●●●○○
Return: ●●●○○
Qualifies for Courage Award: **Yes**

Smart Mop Robot is an artificially intelligent mop and vacuum robot. With its additional mop function not only the dust gets vacuumed, but also persistent stains are removed at the same time.

Smart Mop Robot also has an innovative **mapping function** through which the robot can be sent to any place via an app.

Due to operating in an unknown market, it is possible that there will be **no market demand at all** for Smart Mop Robot. However, if the company succeeds in convincing customers of this innovation, it could generate **high sales and a large market share** as a first mover.

The initial investment to start the Smart Mop project amounts to **3m Lira**.

Financials

The current balance of the project account is 5m Lira.

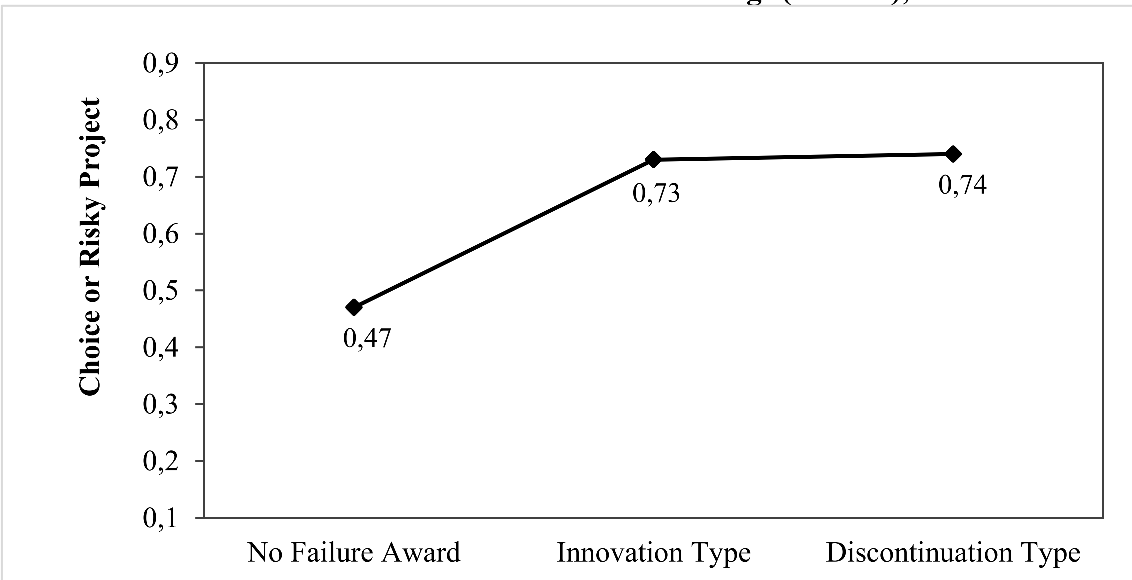
The following **predicted values** were calculated based on currently known information:

	Best-Case Scenario	**Worst-Case Scenario**
Probability	**50%**	**50%**
Required initial investment	3m Lira	3m Lira
Expected project cash inflows	14m Lira	0m Lira

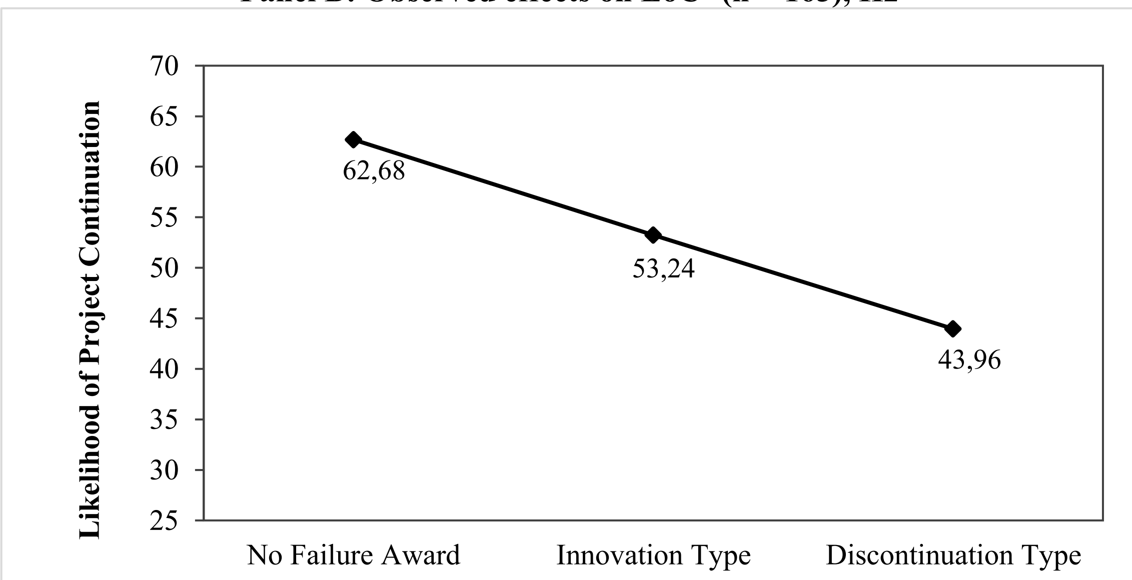
Figure 5

Observed Effects of Failure Awards on Risk-Taking for all Participants (H1) and on EoC for Participants Choosing the Riskier Project (H2)

Panel A: Observed Effects on Risk-Taking[a] (n = 264), H1



Panel B: Observed effects on EoC[b] (n = 165), H2



[a] The dependent variable risk-taking is approximated by the participants' choice of the riskier project, which is a binary variable with 0 = choice of the safe project and 1 = choice of the riskier project. The riskier project is the Smart Mop Robot with an expected value of 7 m and a variance of 49 (SD of 7.0), and the safe project is the Smart Vacuum Robot with an expected value of 6.5 m and a variance of 2.25 (SD of 1.5)

[b] The dependent variable Escalation of Commitment is approximated by the participants' recommendation to continue a poorly performing project, measured on a 101-scale (0 = termination, 100 = continuation). We manipulate the type of the Failure Award at two levels (innovation vs. discontinuation) and added a Failure Award absent treatment for which no Failure Award is present.

References

Anthony, S. D. (2020). *Eat, Sleep, Innovate: How to Make Creativity an Everyday Habit Inside Your Organization*. La Vergne: Harvard Business Review Press.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational behavior and human decision processes, 35*, 124–140.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological review, 64, Part 1*, 359–372.

Baer, M., & Frese, M. (2003). Innovation is not enough: climates for initiative and psychological safety, process innovations, and firm performance. *Journal of Organizational Behavior, 24*, 45–68.

Barki, H., Rivard, S., & Talbot, J. (1993). Toward an Assessment of Software Development Risk. *Journal of Management Information Systems, 10*(2), 203–225.

Barton, S. L., Duchon, D., & Dunegan, K. J. (1989). An Empirical Test of Staw and Ross's Prescriptions for the Management of Escalation of Commitment Behavior in Organizations. *Decision Sciences, 20*, 532–544.

Baysinger, B. D., Kosnik, R. D., & Turk, T. A. (1991). Effects of Board and Ownership Structure on Corporate R&D Strategy. *Academy of Management Journal, 34*, 205–214.

Behrens, J., & Ernst, H. (2014). What Keeps Managers Away from a Losing Course of Action? Go/Stop Decisions in New Product Development. *Journal of Product Innovation Management, 31*, 361–374.

Bentley, J. W. (2021). Improving the Statistical Power and Reliability of Research Using Amazon Mechanical Turk. *Accounting Horizons*.

Birkinshaw, J., & Haas, M. (2016a). Fail more, fail better. *London Business School Review, 27*, 34–37.

Birkinshaw, J., & Haas, M. (2016b). Increase Your Return on Failure. *Harvard Business Review*, 88–93.

Brasel, K., Doxey, M., Grenier, J. H., & Reffett, A. (2016). Risk Disclosure Preceding Negative Outcomes: The Effects of Reporting Critical Audit Matters on Judgments of Auditor Liability. *The Accounting Review, 91*, 1345–1362.

Brassey, J., van Dam, N., & van Witteloostuijn, A. (2019). *Advancing Authentic Confidence through Emotional Flexibility: An evidence-based playbook of insights, practices and tools to shape your future*. Lulu.com.

Brink, A. G., Gouldman, A., Rose, J. M., & Rotaru, K. (2020). Effects of superiors' compensation structures on psychophysiological responses and real earnings management decisions of subordinate managers. *Management Accounting Research, 48*, 100691.

Brockner, J. (1992). The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress. *The Academy of Management Review, 17*, 39.

Brockner, J., Rubin, J. Z., & Lang, E. (1981). Face-saving and entrapment. *Journal of Experimental Social Psychology, 17*, 68–79.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on psychological science : a journal of the Association for Psychological Science, 6*, 3–5.

Cannon, M. D., & Edmondson, A. C. (2005). Failing to Learn and Learning to Fail (Intelligently). *Long Range Planning, 38*, 299–319.

Chen, D. L., Schonger, M., & Wickens, C. (2016). *oTree - An Open-Source Platform for Laboratory, Online, and Field Experiments*.

Cheng, M. M., Schulz, A. K.-D., Luckett, P. F., & Booth, P. (2003). The Effects of Hurdle Rates on the Level of Escalation of Commitment in Capital Budgeting. *Behavioral Research in Accounting, 15*, 63–85.

Clifford, C. (2019, December 2). Coke CEO: Why we have an award for projects that fail. *CNBC.*https://www.cnbc.com/2019/12/02/coke-ceo-quincey-why-coca-cola-co-has-award-for-projects-that-fail.html.

Crosetto, P., & Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of Risk and Uncertainty, 47*, 31–65.

Denison, C. A. (2009). Real Options and Escalation of Commitment: A Behavioral Analysis of Capital Investment Decisions. *The Accounting Review, 84*, 133–155.

Deutschman, A. (2004, December 1). Gore's Text for Innovation: W.L. Gore's new rules of business start with breaking the old rules. *Fast Company.* https://www.fastcompany.com/51510/gores-text-innovation.

Druckman, J. N., & McDermott, R. (2008). Emotion and the Framing of Risky Choice. *Political Behavior, 30*, 297–321.

Edmondson, A. C. (1999). Psychological safety and learning behavior in teams. *Administrative Science Quarterly, 44*, 250–282.

Edmondson, A. C. (2003). Managing the Risk of Learning: Psychological Safety in Work Teams. In M. A. West, D. Tjosvold, & K. G. Smith (Eds.), *International Handbook of Organizational Teamwork and Cooperative Working* (pp. 255–275). Chichester, UK: John Wiley & Sons Ltd.

Edmondson, A. C., & Lei, Z. (2014). Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*, 23–43.

Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review, 14*, 57–74.

Farrell, A. M., Grenier, J. H., & Leiby, J. (2017). Scoundrels or Stars? Theory and Evidence on the Quality of Workers in Online Labor Markets. *The Accounting Review, 92*, 93–114.

Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, Calif.: Stanford Univ. Press.

Fischer, S., Frese, M., Mertins, J. C., & Hardt-Gawron, J. V. (2018). The Role of Error Management Culture for Firm and Individual Innovativeness. *Applied Psychology, 67*, 428–453.

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological Safety: A Meta-Analytic Review and Extension. *Personnel Psychology, 70*, 113–165.

Freiberg, K. (2011). *Nanovation: How a Little Car Can Teach the World to Think Big and Act Bold*. Nashville: HarperCollins Leadership.

García-Granero, A., Llopis, Ó., Fernández-Mesa, A., & Alegre, J. (2015). Unraveling the link between managerial risk-taking and innovation: The mediating role of a risk-taking climate. *Journal of Business Research, 68*, 1094–1104.

Giorgi, S. (2017). The Mind and Heart of Resonance: The Role of Cognition and Emotions in Frame Effectiveness. *Journal of Management Studies, 54*, 711–738.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Knopf Doubleday Publishing Group.

Gold, A., Gronewold, U., & Salterio, S. E. (2014). Error Management in Audit Firms: Error Climate, Type, and Originator. *The Accounting Review, 89*, 303–330.

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of experimental psychology. General, 150*, 700–709.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples. *Journal of Behavioral Decision Making, 26*, 213–224.

Grossman, D. (2014, January 24). Secret Google lab 'rewards staff for failure'. *BBC News.* https://www.bbc.com/news/technology-25880738.

Guggenmos, R. D., Piercey, M. D., & Agoglia, C. P. (2018). Custom Contrast Testing: Current Trends and a New Approach. *The Accounting Review, 93*, 223–244.

Harrell, A., & Harrison, P. (1994). An incentive to shirk, privately held information, and managers' project evaluation decisions. *Accounting, Organizations and Society, 19*, 569–577.

Heng, C.-S., Tan, B. C., & Wei, K.-K. (2003). De-escalation of commitment in software projects: Who matters? What matters? *Information & Management, 41*, 99–110.

Hennig, J. C., Ahrens, C., Oehmichen, J., & Wolff, M. (2023). Employee stock ownership and firm exit decisions: A cross-country analysis of rank-and-file employees. *Accounting, Organizations and Society, 104*, 101390.

Highhouse, S., & Yüce, P. (1996). Perspectives, Perceptions, and Risk-Taking Behavior. *Organizational behavior and human decision processes, 65*, 159–167.

Hirshleifer, D. (1993). Managerial Reputation and Corporate Investment Decisions. *Financial Management, 22*, 145.

Hunt, N. C., & Scheetz, A. M. (2019). Using MTurk to Distribute a Survey or Experiment: Methodological Considerations. *Journal of Information Systems, 33*, 43–65.

James, L. R., Hartman, A., Stebbins, M. W., & Jones, A. P. (1977). Relationships between psychological climate and a VIE model for work motivation. *Personnel Psychology, 30*, 229–254.

Johnson, R. (2017, December 15). Why Google rewards its employees for failing. *LinkedIn.* https://www.linkedin.com/pulse/why-google-rewards-its-employees-failing-ron-johnson/.

Kachelmeier, S. J., Thornock, T. A., & Williamson, M. G. (2016). Communicated Values as Informal Controls: Promoting Quality While Undermining Productivity? *Contemporary Accounting Research, 33*, 1411–1434.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal, 33*, 692–724.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica, 47*, 263.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35*, 401–415.

Kark, R., & Carmeli, A. (2009). Alive and creating: the mediating role of vitality and aliveness in the relationship between psychological safety and creative work involvement. *Journal of Organizational Behavior, 30*, 785–804.

Keil, M., Tan, B., Wei, K.-K., Saarinen, T., Tuunainen, V., & Wassenaar, A. (2000). A cross-cultural study on escalation of commitment behavior in software projects. *MIS quarterly = Management information systems quarterly, 24*(2), 299–325.

Kreilkamp, N., Matanovic, S., Sommer, F., & Wöhrmann, A. (2021). The Effect of Compensation Caps on Risk-Taking. *Journal of Management Accounting Research, 33*, 77–95.

Kuvalekar, A., & Ravi, N. (2019). When and How to Reward Bad News. *SSRN Electronic Journal*.

Laux, H., Gillenkirch, R. M., & Schenk-Mathes, H. Y. (2018). *Entscheidungstheorie*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Leber, J. (2016, April 14). How Google's Moonshot X Division Helps Its Employees Embrace Failure. *Fast Company.* https://www.fastcompany.com/3058866/how-googles-moonshot-x-division-helps-its-employees-embrace-failure.

Lerner, J. S., & Keltner, D. (2001). Fear, anger, and risk. *Journal of personality and social psychology, 81*, 146–159.

Liu, Y., Huang, Z., Jiang, L., & Messier, W. F. (2020). Are Investors Warned by Disclosure of Conflicts of Interest? The Moderating Effect of Investment Horizon. *The Accounting Review, 95*, 291–310.

Mahlendorf, M. D. (2015). Allowance for failure: reducing dysfunctional behavior by innovating accountability practices. *Journal of Management & Governance, 19*, 655–686.

Malmendier, U., & Tate, G. (2005). CEO Overconfidence and Corporate Investment. *The Journal of Finance, 60*(6), 2661–2700.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological review, 115*, 502–517.

Morgan, J. (2015, March 30). Why Failure Is The Best Competitive Advantage. *Forbes.* https://www.forbes.com/sites/jacobmorgan/2015/03/30/why-failure-is-the-best-competitive-advantage/?sh=3c86f75259df.

NASA. (2021). NASAPeople. https://nasapeople.nasa.gov/awards/award_criteria.htm.

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. *Human Resource Management Review, 27*, 521–535.

Palanski, M. E., & Vogelgesang, G. R. (2011). Virtuous Creativity: The Effects of Leader Behavioural Integrity on Follower Creative Thinking and Risk Taking. *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration, 28*, 259–269.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running Experiments Using Amazon Mechanical Turk. *Judgment and Decision Making, 5*(5), 411–419.

Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology, 70*, 153–163.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*, 1023–1031.

Proulx, M. (2019). Why We Give Out an 'Epic Fail Award' to Employees. Innovation Leader. https://www.innovationleader.com/employee-engagement/why-we-give-out-an-epic-fail-award-to-employees/1193.article.

Roetzel, P. G., Pedell, B., & Groninger, D. (2020). Information load in escalation situations: combustive agent or counteractive measure? *Journal of Business Economics, 90*, 757–786.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2011). *Contrasts and Effect Sizes in Behavioral Research*. Cambridge University Press.

Schoemaker, P. J. H. (1982). The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature, 20*(2), 529–563.

Schoorman, F. D., & Holahan, P. J. (1996). Psychological antecedents of escalation behavior: effects of choice, responsibility, and decision consequences. *Journal of Applied Psychology, 81*, 786–794.

Seybert, N. (2010). R&D Capitalization and Reputation-Driven Real Earnings Management (Partially Retracted). *The Accounting Review, 85*, 671–693.

Simonson, I., & Staw, B. M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *The Journal of applied psychology, 77*, 419–426.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the Determinants of Risk Behavior. *Academy of Management Review, 17*, 9.

Sitkin, S. B., & Weingart, L. R. (1995). Determinants of Risky Decision-Making Behavior: A Test of the Mediating Role of Risk Perceptions and Propensity. *Academy of Management Journal, 38*, 1573–1592.

Sleesman, D. J., Conlon, D. E., McNamara, G., & Miles, J. E. (2012). Cleaning Up the Big Muddy: A Meta-Analytic Review of the Determinants of Escalation of Commitment. *Academy of Management Journal, 55*, 541–562.

Sprinkle, G. B., Williamson, M. G., & Upton, D. R. (2008). The effort and risk-taking effects of budget-based contracts. *Accounting Organizations and Society, 33*, 436–452.

Staw, B. M. (1976). Knee-deep in the big muddy: a study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance, 16*, 27–44.

Staw, B. M., & Ross, J. (1987). Behavior in escalation situations: Antecedents, prototypes, and solutions. *Research in Organizational Behavior, 9*, 39–78.

Staw, B. M., & Ross, J. (1989). Understanding behavior in escalation situations. *Science (New York, N.Y.), 246*, 216–220.

Steinkühler, D., Mahlendorf, M. D., & Brettel, M. (2014). How Self-Justification Indirectly Drives Escalation of Commitment. *Schmalenbach Business Review, 66*, 191–222.

Stewart, H. (2015). 8 companies that celebrate mistakes. https://www.happy.co.uk/blogs/8-companies-that-celebrate-mistakes/.

Supercell. (2021). Our story × Supercell. https://supercell.com/en/our-story/.

TATA. (2021). Tata Innovista. https://tatainnovista.com/.

Tsai, M.-H., & Young, M. J. (2010). Anger, fear, and escalation of commitment. *Cognition & Emotion, 24*, 962–973.

van Dyck, C., Frese, M., Baer, M., & Sonnentag, S. (2005). Organizational error management culture and its impact on performance: a two-study replication. *The Journal of applied psychology, 90*, 1228–1240.

Waczek, A. (2012, September 18). Dare to Try. *Epic Work Epic Life.* https://epicworkepiclife.com/dare-to-try-2/. Accessed 21 December 2021.

Wiseman, R. M., & Gomez-Mejia, L. R. (1998). A Behavioral Agency Model of Managerial Risk Taking. *Academy of Management Review, 23*, 133–153.

Wong, K. F. E. (2005). The Role of Risk in Making Decisions under Escalation Situations. *Applied Psychology, 54*, 584–607.

Wong, K. F. E., Yik, M., & Kwong, J. Y. Y. (2006). Understanding the emotional aspects of escalation of commitment: the role of negative affect. *Journal of Applied Psychology, 91*, 282–297.

Wu, H.-L. (2008). When does internal governance make firms innovative? *Journal of Business Research, 61*, 141–153.

Zhou, J., & George, J. M. (2001). When Job Dissatisfaction Leads to Creativity: Encouraging the Expression of Voice. *Academy of Management Journal, 44*, 682–696.

Zhou, Q., & Pan, W. (2015). A Cross-Level Examination of the Process Linking Transformational Leadership and Creativity: The Role of Psychological Safety Climate. *Human Performance, 28*, 405–424.

What Motivates Independent Directors? The Influence of Director Incentives on Director Decisions

Authors Corinna Ewelt-Knauer, Hannes Gerstel, Mohamed Khaled, Arnt Wöhrmann

Abstract Reputation, risk, workload, and monetary incentives affect independent directors' decisions to join or leave board of directors. Using relative incentive proxies, we find that directors strategically relinquish board seats to increase their reputation. Changing the perspective, we hypothesize and show that accepting an additional board nomination is incentivized by a director's goal to increase reputation growth. Lastly, by taking on a firm perspective, we find a positive association between the board's average portfolio reputation and the risk of all outside directors on firm performance and earnings management. Our findings have practical relevance for firms wanting to retain skilled directors and for firms trying to attenuate earnings management. Finally, our findings also have practical implications for directors considering whether to accept an additional board nomination, as accepting is generally beneficial due to future increases in reputational capital.

Publication Status Revise & Resubmit

Managerial and Decision Economics (VHB-Jourqual 3: B)

Workshop/ Conference Presentations 2022 CAAA Annual Meeting (Saskatoon)

2022 BFGA (Giessen)

Acknowledgements We thank Friedrich Sommer and Patricia O'Malley for their helpful comments and suggestions.

Author Contribution	Ewelt-Knauer	Gerstel	Khaled	Wöhrmann
Relative Share	*5%*	*45%*	*45%*	*5%*
Conceptualization	✓	✓	✓	✓
Theory Development		✓	✓	
Methodology		✓	✓	
Data Acquisition		✓	✓	
Data Analyses		✓	✓	
Writing		✓	✓	

1 Introduction

The introduction of the Sarbanes-Oxley Act (SOX) in 2002 reshaped the labor market for independent directors[1] substantially by mandating firms to more frequent board meetings and requiring a minimum level of independence for their board committees (Linck et al. 2009; Chen and Moers 2018). These new corporate governance mandates increased demand for outside directors, resulting in many outside directors occupying multiple board seats. However, it is unclear how directors manage their portfolios of all existing and potential new directorships to deal with the challenges of dynamic labor markets.

For example, the current CEO of Microsoft (Satya Nadella) joined Riverbed Technology as an outside director in 2012 after already holding an outside directorship at BravoSolution US. After becoming Microsoft's CEO in 2013, he relinquished his seat at Riverbed in 2014. Prior research identifies four major incentives –compensation, risk, workload, and reputation– influencing directors to join or leave a board (e.g., Boivie et al. 2012; Masulis and Mobbs 2014; Ormazabal 2018). Since outside director compensation is marginal compared to their executive compensation (Adams and Ferreira 2008), one could argue that he left Riverbed due to time constraints. However, two years later, he joined Starbucks as an outside director. Our research tries to shed light on why outside directors relinquish board seats (H1) and whether gaining an additional board seat is advantageous for a director's reputation growth compared to only holding one board seat (H2). Lastly, we examine the effect of average portfolio characteristics of all outside directors serving on a board on firm outcomes (H3).

First, we analyze how the different director incentives impact their decisions to relinquish a directorship. (e.g., Satya Nadella leaving Riverbed Technology). However, it remains

[1] Independent directors (also defined as outside directors) assume a primarily role as monitors and advisors of firms and are nominally independent from the executive management (Boyd 1990). In the following, we use the word *directors* to refer to independent directors.

to be determined whether and to what extent compensation, risk, workload, and reputational incentives influence the composition of outside directorship portfolios. Based on the reputation hypothesis by Fama and Jensen (1983), we argue that relative reputational concerns constitute the dominant incentive for outside directors confronted with the decision to retain or depart from an existing directorship. Thus, we posit that an outside director is likelier to relinquish a directorship that ranks relatively lower than her other directorships (H1).

To measure the impact of relative director incentives on their relinquish decisions, we calculate the four relative incentive measures by comparing a director's firm to all other firms of her directorship portfolio. Using a director-firm-year perspective and two-way fixed effects linear regressions, we find that directors are more likely to relinquish a directorship if the respective firm has a relatively lower reputation than the rest of her portfolio and when it requires relatively more working hours.

Second, in H1, we find that directors strategically relinquish board seats to increase their reputation. Changing the perspective, we argue that accepting an additional board nomination is incentivized by a director's goal to grow her reputation (e.g., Satya Nadella accepting Riverbed as his second directorship). Thus, we investigate whether gaining an additional seat positively impacts directors' reputation growth compared to directors who only hold one outside directorship (H2). In the case of receiving and accepting an additional board nomination, we assume that a potential negative impact of a decrease in monitoring efficacy due to increased workload (Bar-Hava et al. 2020) is outweighed by two positive aspects: (1) nomination committees tend to select more successful directors in the first place (Booth and Deli 1996; Brickley et al. 1999), and (2) directors could gain and share additional information and expertise by holding multiple directorships, which can subsequently be used to increase monitoring efficacy.

By applying a generalized difference-in-difference design and implementing several robustness checks (e.g., generating pseudo-events and applying propensity score matching), we find that holding multiple outside directorships significantly increases the director's reputational growth compared to directors who only hold one board seat. Additional analyses reveal that the increased reputation growth could be driven by the information and knowledge-sharing synergies since leaving an additional board seat leads to lower reputation growth afterward.

Third, taking on a firm perspective, it is unclear how specific director portfolio compositions (e.g., a board of directors' average reputation and risk propensity) affect firm outcomes. E.g., does the composition of Starbucks' board of directors impact its firm performance? For H3a, we argue that directors with higher reputational capital are more incentivized to protect their reputation and thus monitor more effectively, leading to increases in firm performance. Next, we analyze whether outside directors' risk-taking propensity is related to firms' earnings management (H3b). Even though outside directors monitoring activities should decrease litigation risks, prior research finds mixed evidence regarding the proportion of outside directors on boards and decreases in earnings management (e.g., Wang et al. 2015; Badolato et al. 2014; Klein 2002). To shed light on these mixed findings, we argue that outside directors' risk-taking propensity is related to firms' earnings management. Consequently, we propose that outside directors holding more risky directorships in their directorship portfolio are more inclined to accept earnings management and its related risks.

Taking the firm perspective and applying two-way fixed-effects linear regressions, we find that a higher average portfolio reputation of independent board members is associated with better firm performance (H3a). Furthermore, we observe that a higher average risk level in the portfolios of outside directors is positively associated with increases in earnings management (H3b). As additional analyses, we replace the absolute average incentive measures with a relative average measure reflecting the proportion of directors on a board for whom the directorship is relatively higher ranked than her other directorships (similar to Sila et al. 2017), confirming the findings of H3a.

Our paper contributes in several ways to theory and practice. First, we investigate if relative reputational concerns govern the composition of directorship portfolios. We use relative instead of absolute incentives since Ormazabal (2018) notes that investigations into the relative attributes of directorships are still marginal in the current empirical literature. In this vein, we simultaneously analyze the impact of four major director incentives (compensation, risk, workload, and reputation). Contrary to previous studies (Ormazabal 2018; Masulis and Mobbs 2014), we do not limit our investigation to the highest (lowest) ranked directorship. Especially if a director holds more than two seats (3,817 directors in our sample), then information on the other directorships is ignored. Instead, we argue that reputational incentives should be considered in the context of all other directorships in a directorship portfolio. Hence, we expand the scope of the investigation to include all directorships of a director's portfolio for our relative incentive measures.

Our results have practical relevance for firms wanting to retain directors. By showing that relative workload and reputation incentives are highly relevant when directors contemplate relinquishing an existing directorship, firms should manage their directors' (relative) workload. Furthermore, firms should consider the director's relative portfolio reputation when offering a seat.

Second, to our best knowledge, prior literature focuses on determinants leading to relinquish decisions (e.g., Boivie et al. 2012; Ormazabal 2018; Masulis and Mobbs 2014). We add to the literature by examining the effects of accepting additional directorships on directors' reputation growth. Our results show that holding multiple outside directorships significantly increases the director's reputational growth compared to directors who only hold one board seat. These findings have practical implications for directors contemplating accepting an additional nomination. By accepting an additional nomination, directors can increase their reputation growth, which also could lead to better executive positions.

Lastly, Sila et al. (2017) investigate outside directors' reputational incentives from a firm perspective by examining how the informativeness of stock prices is influenced by the relative reputational ranking of the firm by its outside directors. We expand this research by examining the impact of the board of directors' absolute and relative reputation and risk propensity on key firm-specific characteristics like performance and earnings management. These results have practical relevance for firms. First, we show that hiring reputable directors increases firm performance. Second, firms concerned about heightened earnings management levels should consider evaluating the average firm risk that potential new outside directors are tolerating in their directorship portfolios. Our results also shed light on prior mixed findings concerning the relationship between the proportion of outside directors and firms' earnings management (e.g., Wang et al. 2015; Klein 2002) since we show that the board's portfolio risk increases earnings management.

In conclusion, we expand the still marginal literature on the effect of (relative) outside director incentives on the composition of directorship portfolios and the consequences of the directorship portfolio attributes on firm outcomes. The remainder of the paper is structured as follows. Section 2 describes the derivation of our hypotheses. Section 3 presents our data sample. Sections 4 to 7 present the results, and concluding remarks are provided in Section 7.

2 Hypotheses Development

2.1 H1 - Director incentives and relinquish decisions

Acquiring additional outside directorships increases labor market opportunities since holding multiple board seats generally signals a director's competence (Peyer and Perry 2005). However, the number of directorships is restricted by directors' time and effort limitations.

Thus, outside directors are regularly confronted with how to allocate their time and effort efficiently among their existing directorships to fulfill their monitoring duties (Ferris et al. 2003; Hossain and Oon 2022). Ghannam et al. (2019) argue that outside directors evaluate potential new and existing directorships based on a portfolio of incentives. Prior research identifies four major incentives –compensation, risk, workload, and reputation– influencing directors' decision to join or leave a board (e.g., Boivie et al. 2012; Masulis and Mobbs 2014; Ormazabal 2018).

Outside directors might structure the composition of their directorship portfolios based on maximizing their financial benefits (Yermack 2004). However, Adams and Ferreira (2008) find that outside director compensation represents an arguably small fraction of directors' total wealth. Regarding the risk incentive, Ormazabal (2018) provides evidence that risk might be another important determinant for a director to depart or acquire an additional board seat. Specifically, Ormazabal (2018) finds a higher likelihood of outside directors leaving their riskiest directorship after the financial crisis to avoid reputational repercussions due to negative firm events. Next, Fahlenbrach et al. (2010) provide evidence that outside directors are more likely to depart from a directorship in anticipation of events that would substantially increase the workload necessary to fulfill their board duties. Lastly, prior literature considers reputational benefits as a central incentive for outside directors to serve on boards and acquire additional directorships of prestigious firms (e.g., Masulis and Mobbs 2014; Fich 2005). Serving on the board of larger firms constitutes a signal of a director's competence and leads to higher visibility in the labor market (Fama and Jensen 1983).

Thereby, directors can expand their career opportunities as inside and outside directors (Shivdasani 1993). Yermack (2004) and Fich (2005) support this thought by noting that outside directors of larger firms are more likely to gain additional directorships. Cowen and Marcel (2011) expand on this research and find that directors often relinquish their seats in distressed firms, e.g., after lawsuits or restatements occur. Moreover, Masulis and Mobbs (2014) find that directors are more likely to resign from poorly performing firms with a relatively low reputation. Concerning the interplay of reputation and workload incentives, they also show that directors exhibit a lower absence rate at board meetings when the firm has a relatively high reputation in a director's portfolio.

We argue that the other three incentives –compensation, risk, and workload– are also (implicitly) affected by reputational concerns. First, by serving on boards of prestigious firms to showcase one's abilities, directors can expand their career opportunities, positively influencing future monetary compensation (Mobbs 2013). Second, directors may depart from riskier firms to protect their reputational capital as adverse circumstances (e.g., financial fraud or poor firm performance) might cast doubt regarding the director's monitoring performance (Boivie et al. 2012). Third, directors potentially want to limit the workload of their directorships since overly busy directors are assessed as less capable regarding monitoring and risk oversight (Fich and Shivdasani 2004), decreasing a director's reputation. In conclusion, based on the reputation hypothesis by Fama and Jensen (1983), we argue that relative reputational concerns constitute the dominant incentive for outside directors confronted with the decision to retain or depart from an existing directorship. We state H1 as follows:

***H1:* Outside directors are more likely to relinquish a directorship that possesses a relatively lower reputation compared to the average reputational value of the directorship portfolio.**

2.2 H2 – Director portfolio adjustments and reputation growth

Nearly half of outside directors within S&P 1500 firms only serve on one board of directors (Masulis and Mobbs 2014). On the contrary, the other half of outside directors change the composition of their directorship portfolio by joining new boards and leaving existing boards. Director exits are either voluntary or forced (Boivie et al. 2012). In contrast to exits, where the underlying reasons behind the exit are usually private (Bar-Hava et al. 2020), directors' decisions to join boards are voluntary and observable. Gaining an additional outside directorship is based on accepting a nomination from the nominating committee and thus represents a voluntary decision by the joining director. In H1, we hypothesize that directors strategically relinquish board seats to increase their reputation. Similarly, we argue that accepting an additional board nomination is incentivized by a director's goal to grow her reputation.

We posit that gaining an additional board seat increases directors' reputation growth for two reasons. First, an outside director needs to be considered and elected by a firm's nomination committee to get the possibility of gaining an additional seat (Callahan et al. 2003; Duchin et al. 2010). Nomination committees could either consider candidates who do not hold an active outside directorship position. Alternatively, they could nominate a director already holding at least one other outside directorship. By being active on another board, directors signal prior monitoring experience that nominating committees can potentially observe. Therefore, we argue that successful outside directors are more likely to get the opportunity to join additional boards in the first place (Brickley et al. 1999). In this vein, Shiah-Hou and Cheng (2012) find that firms with outside directors having more work experience have a higher market performance. Furthermore, firms with directors holding multiple board appointments generally perform better (Booth and Deli 1996).

Second, we argue that serving on multiple boards leads to information synergies and increases in the industry- and firm-specific expertise of the respective outside director (Masulis 2020), potentially increasing the effectiveness of the outside director's monitoring abilities. Moreover, sharing this information and expertise could lead to a positive spillover to other board members. Both factors would positively affect a firm's growth potential, resulting in higher reputation growth for its directors. Corroborating this reasoning, Kor and Sundaramurthy (2009) find that increased industry- and firm-specific experience of outside directors increases firms' sales growth. In this vein, we posit that directors with only one outside directorship most likely miss out on the opportunity to gain and share additional information and expertise from serving on additional boards.

Nevertheless, serving on multiple boards could also have disadvantages for outside directors. By accepting additional directorships, directors limit their attention to each directorship (Bar-Hava et al. 2020). Consequently, directors with multiple directorships might put effort into retaining them instead of closely monitoring the firm's executives (Mace 1986). Corroborating this finding, Balsmeier et al. (2015) show that monitoring tends to decrease with the number of directors who hold multiple directorships. On the contrary, Masulis and Mobbs (2014) posit that outside directors have strong incentives to be viewed as careful monitors by the external labor market, irrespective of their number of board seats. Overall, prior research finds mixed evidence on the relationship between director busyness and firm performance (e.g., Ferris et al. 2003; Fich and Shivdasani 2004; Bar-Hava et al. 2020).

Finally, we assume that a potential negative impact of decreased monitoring efficacy due to increased workload is outweighed by nomination boards selecting more successful directors and information and knowledge sharing synergies from holding multiple directorships. We state H2 as follows:

H2: **Outside directors, who accept at least one additional outside directorship, achieve higher reputation growth than outside directors who only hold one directorship.**

2.3 H3 – Outside director incentives and firm outcomes

A public firm's board's primary role is to protect shareholder interests by monitoring and advising the management regarding corporate decision-making processes (Shiah-Hou and Cheng 2012). Monitoring the management is crucial for firms that separate ownership and control to prohibit potential agency conflicts. As managers are motivated to maximize their personal utility (e.g., their compensation), this potentially leads to management decisions that might contradict the shareholders' best interest (Jensen and Meckling 1976). In this context, the characteristics of outside directors become increasingly important in explaining effective monitoring behavior (Ghannam et al. 2019). For example, Masulis et al. (2012) provide empirical evidence that demographic factors influence the monitoring performance of outside directors. They find that a higher share of foreign outside directors is associated with decreased financial reporting quality and firm performance. In contrast, more foreign directors increase acquisition performance when foreign targets are acquired (Masulis et al. 2012).

Sila et al. (2017) link portfolio-based reputational incentives to firm-specific outcomes. They find that a firm's voluntary disclosure and stock price informativeness increase if its directors rank the firm reputationally higher than their other directorships. They argue that outside directors try to protect their reputation, which leads to heightened demand for reliable public information and greater distrust of private information provided by the management. Fredriksson et al. (2020) support this notion by finding that high-reputation directors enforce better audit quality in their directorships to protect their reputational capital. Similarly, we argue that directors with higher reputational capital are more incentivized to protect their reputation and thus monitor more effectively, leading to increases in firm performance. Fich's (2005) findings support this notion by identifying a more positive share price reaction to announcements of high-reputation director appointments.

Moreover, prior research notes that high-reputation directors are rewarded for effective monitoring by increased labor market opportunities and can therefore acquire additional directorships (Fama and Jensen 1983; Fich and Shivdasani 2007). In this regard, the higher reputational capital of outside directors can be understood as a signal for their monitoring ability and experience. Similarly, outside directors with higher reputational capital might also be more trusted advisors and possess a greater influence on the management of a firm. Both factors lead to increased firm performance monitored by high-reputation boards.

Another firm outcome influenced by directors' monitoring activities is accrual-based earnings management. Earnings management via abnormal accruals can be used to alter financial reports regarding earnings (Beneish 2001). For example, income-increasing earnings management might be facilitated to obscure investors' perception of the firm's economic situation, and executives can initiate income-increasing earnings management to maximize their performance-based compensation (Healy and Wahlen 1999). As a negative effect, engaging in earnings management significantly increases litigation risks (e.g., Ibrahim et al. 2011; Palmrose et al. 2004; Lo 2008). Even though outside directors' monitoring activities should decrease litigation risks, prior research finds mixed evidence regarding the proportion of outside directors on boards and decreases in earnings management (e.g., Badolato et al. 2014; Klein 2002; Wang et al. 2015).

To shed light on these mixed findings, we argue that outside directors' risk-taking propensity is related to firms' earnings management. In this vein, Deng et al. (2018) show that risk-seeking executives are more likely to utilize earnings management. We assume that negative firm outcomes (e.g., increases in litigation risk) and positive firm outcomes (e.g., reporting higher earnings) affect all board members associated with the firm. Therefore, the propensity to engage in earnings management is not just influenced by top-management positions but by all board members, including outside directors. Fredriksson et al. (2020) posit that less risk-averse outside directors might be incentivized to tolerate more earnings management. We argue that the overall risk outside directors accept in their directorship portfolio is related to the level of earnings management a director is ready to tolerate. Consequently, we propose that outside directors holding more risky directorships in their directorship portfolio are more inclined to accept earnings management and its related risks.

In conclusion, we argue that firm performance is positively associated with higher levels of average reputation of its outside directors since it is a signal for their monitoring ability and experience (H3a). For H3b, we posit that firms' accrual-based earnings management increases with their outside directors' risk propensity since they are more inclined to accept the related risks of earnings management. We state the hypotheses as follows:

***H3a:* Higher overall reputation levels in the directorship portfolios of outside directors on a firm's board are positively associated with firm performance.**

***H3b:* Higher overall levels of risk-bearing in the directorship portfolios of outside directors on a firm's board are positively associated with accrual-based earnings management.**

3 Sample Selection and Data Description

Our sample contains 22,460 outside directors serving on boards of 11,347 firms between 1999 and 2019. Director data is obtained from the database BoardEx, which provides biographical and relationship data on the boards of public and private companies starting in the year 1999. Our sample is restricted to public companies with relevant accounting data available. Our investigation only pertains to the directorship portfolios of outside directors, as inside and grey directors generally possess different reputational incentives that obstruct a direct comparison (Masulis and Mobbs 2014). Therefore, we restrict the sample to outside director observations by excluding observations with executive directors. Firm-specific data is obtained from Thomson Reuters Refinitiv.

For H1 (directors relinquish decisions), we use a director-firm-year perspective with 32,970 observations, covering only directors who held at least two board seats during our observation period (similar to Masulis and Mobbs 2014). To examine the impact of join decisions on directors' reputation growth (H2), we use a director-year perspective with 31,283 observations. Lastly, we use a firm-year perspective with 8,289 observations to analyze the impact of board incentives on firm outcomes (H3).

4 Tests of H1 – Director-Firm-Level Analyses: Director Reputational Concerns and Relinquish Decisions

4.1 Identification strategy

To identify the effect of a director seat's relative reputation on the decision to relinquish a directorship, we apply the following two-way fixed effects regression models:

$$Relinquished_{i,t} = \alpha + \beta_1 RelativeReputation_{i,t} + \sum \beta_k Controls_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t} \quad (1)$$

where i indexes the director-firm unit and t indexes the year. Our dependent variable is relinquished, which is an indicator variable that equals one if the outside director leaves a firm's board in a given year, and zero if she has an active position in the firm's board (similar to Ormazabal 2018; Masulis and Mobbs 2014).[2]

Following prior research (e.g., Masulis and Mobbs 2014; Ormazabal 2018; Bryan and Mason 2020; Sila et al. 2017), we use market capitalization as our proxy for firm reputation. Our main independent variable *–ReputationDummy–* is fundamentally based on Masulis and Mobbs (2014). They classify a directorship as high (low) ranked if the firm is 10% larger (smaller) than the director's smallest (largest) directorship, based on market capitalization. The reputation measurement based on being larger (smaller) than the smallest (largest) directorship has the disadvantage of losing relevant information. Especially if a director holds more than two seats (3,817 directors in our sample), then information of the other directorships is ignored. For example, suppose a director has a portfolio with four board seats, of which the second and

[2] BoardEx reports the exact join and end date of a directorship for most observations. Since our models use yearly observations, we round start and end dates based on the actual reported month. If the reported month lies between July and December, then we set the date to the 31st of December of the same year. If the reported month lies between January and June, then we set the date backwards to the 31st of December of the previous year. The other option of always setting the reported date to the 31st of December of the given year would lead to larger offsets for reported dates in the first half of the year. Consequently, shifting the reported date up to +/- 6 months reduces the potential offset. In case of missing exact dates, we set the 31st of December as the date.

third firms (ranked in market cap) deviate significantly from the smallest and largest observation. In that case, the second and third directorships are ignored when calculating a high and low-ranked indicator variable.

Therefore, we include all firms of a director's portfolio in our *ReputationDummy* measurement. The dummy variable equals one if the firm's market capitalization is larger than the average market capitalization of all other firms in the portfolio, and zero if it is smaller (or equal). Thereby, we explicitly consider all firms of a directorship portfolio instead of just comparing with the most (least) reputable firm. Analogously, we calculate *RelativeReputation* as an additional measure to the dummy variable, where the market capitalization of a firm i in year t of director j is divided by the average market capitalization of the remaining firms in the director portfolio:

$$RelativeReputation = MarketCap_{i,t} \Big/ \frac{(Total\ Market\ Cap_{j,t} - Market\ Cap_{i,t})}{(Directorships_{j,t} - 1)} \quad (2)$$

To reduce potential endogeneity concerns, we implement director-level fixed effects (γ_i) controlling for time-invariant director characteristics (similar to Ormazabal (2018)). Additionally, we apply industry-year-level fixed effects (δ_t) to control for industry shocks affecting the probability of a director's exit from a given industry in a given year. Controls is a vector of director, director-firm and firm controls. Standard errors are robust to heteroskedasticity and serial correlations within industry-year level clusters.

4.2 Control variables

Similar to Yermack (2004), Masulis and Mobbs (2014), and Ormazabal (2018), we include several control variables in our regression models, which potentially influence director turnover. We capture director-specific turnover determinants, including director-portfolio (e.g., relative workload) and director controls (e.g., age and number of directorships held), director-firm-specific controls (e.g., tenure of the director in the firm) and firm-specific controls (e.g., size and return on assets):[3]

i. Director-Portfolio-Controls

Ormazabal (2018) finds that (inside and outside) directors tend to relinquish their riskiest directorships in the years after the financial crisis. Therefore, we add firms' beta as a proxy for firm *risk*. Boivie et al. (2012) and Masulis and Mobbs (2014) find that increased director workload (approximated by board meetings) and director compensation (approximated by average board compensation) increase the likelihood of a director exit. Thus, we add *workload* and *compensation* to our portfolio incentive controls.

ii. Director-Controls

We control for potential director retirements by including the natural logarithm of their age (Ormazabal 2018). Further, we control appointments to another firm's board of directors since appointments could impact their decision to leave an existing board seat (Linck et al. 2008). We also control the number of directorships held since leaving a seat has a different impact depending on the total composition of directorships.

iii. Director-Firm-Controls

We include the director's tenure in the firm since a higher tenure makes retirement more likely. Being an active member or the chairman of a committee could reduce the likelihood of leaving a directorship (Yermack 2004). Therefore, we include whether the director was active in the compensation, audit, finance, governance, or risk committee and whether she was a chairman in at least one of the committees. The different committees are listed separately due to their different potential impacts on leaving a board. For example, Masulis and Mobbs (2014) state that the audit and compensation committee are regarded as more time-consuming.

[3] For a detailed list and description see the variable definitions in Appendix A.

iv. Firm-Controls

Since the firm's absolute size is a potential proxy for the firm's reputation and thus might influence a director's decision to leave a board seat, we include the absolute market capitalization to control for any absolute reputation and size impact of the firm.[4] Prior research shows that poor performance increases director turnover (Ormazabal 2018; Yermack 2004). Hence, we include ROA, total asset growth, and Tobin's q as further firm controls. Moreover, since director turnover decreases for firms with a smaller board of directors (Masulis and Mobbs 2014), we include a firm's board size as a control variable. Lastly, we winsorize firm variables at the 1st and 99th percentile (except for variables bound to the 0 – 1 range) to mitigate the impact of outliers.

4.3 Descriptive statistics of the director-firm perspective

Table 1 shows the descriptive statistics for our model concerning director-firm-years, including 32,970 observations.

[Insert Table 1 about here]

Our dummy variable *Relinquished* captures the loss of directorships and indicates that around 6.5% of all director-firm-year observations are related to a loss of a director position. This is in line with Ormazabal (2018), who finds that around 8% of observations include a loss event. Our reputation dummy (*ReputationDummy*) indicates that in 55% of the observations, the firm possesses a larger market capitalization than the remaining directorship portfolio. Analogously, the means of compensation (*CompensationDummy*) and the number of board meetings (*MeetingsDummy*) are also both in the 50% range (0.424 and 0.416). Since the firm beta dummy has a lower standard deviation than the other incentives, the mean of the *RiskDummy* is 0.255. The age of the directors in our sample (*Age*) ranges from 31 years to 94

[4] We use the natural logarithm of the market capitalization to reduce the effect of outliers and to decrease the magnitude of its coefficient.

years, with a mean of 62 years. On average, a director holds 2.8 directorships. The average *tenure* is 5.7 years, while the average firm size equals $20.5 bn (*Size*). On average, firms yield a mean yearly return on assets of 4.1% (*ROA*) and an asset growth rate of 10.4% (*Growth*). Lastly, the firm boards display a mean yearly attendance rate of 82% (*Board_Attendance*).

4.4 Results of H1

For H1, we predict that directors are more likely to relinquish an existing board position if the respective board seat exhibits a lower reputation than the other firms in the director's portfolio. Thus, the impact of the *ReputationDummy* on *Relinquished* (β_1) should be negative and significant.

[Insert Table 2 about here]

Results of Table 2 Model 1 reveal a significant negative coefficient ($t = -3.12$, $p < 0.01$) for our independent variable *ReputationDummy*. This indicates that directors pursue a portfolio approach by considering the relative reputation of their directorships when making relinquish decisions. Thus, they are more likely to relinquish directorships that provide a lower share of their overall reputational capital than the other directorships.

Moreover, we find a strong positive association between the relative number of board meetings (*MeetingsDummy*) and the likelihood of relinquishing a directorship ($t = 3.54$, $p = 0.013$). As the director's time and effort are limited, outside directors that serve on multiple boards might resign from their more workload-intensive director duties. Regarding the risk (*RiskDummy*) and compensation (*CompensationDummy*) incentives, both coefficients are in-significant ($t = 0.11$, $p = 0.910$; $t = -1.21$, $p = 0.228$), indicating that reputation and workload incentives are most relevant when directors decide to relinquish a seat.

4.5 Robustness checks of H1

Lastly, we provide two robustness checks to verify the results of H1. Our base model includes the relative director incentives as dummy variables. To factor in the relative difference in each of the four incentive proxies, we substitute the four dummies with continuous variables capturing the relative difference of a directorship compared to the average of the remaining portfolio for the respective incentive (see Equation 2). Again, Model 2 also displays a highly significant and negative coefficient for our variable that captures *RelativeReputation (*t = -2.69, p < 0.01) and a highly significant and positive coefficient regarding our measure for relative workload differences (t = 4.68, p < 0.01).

Second, there might be a time-lag between directors' relinquishment decisions and the actual relinquishment. It is likely that a director will not immediately relinquish her seat when announcing her exit from the board of directors. Therefore, we lag the control variables by one year.[5] The significance levels of the four incentive proxies remain unchanged for the incentive dummies (Model 3) and the relative incentives (Model 4), with the reputation and workload proxies being significant at the 1% level and the risk and compensation incentives being insignificant. In conclusion, these findings demonstrate that relative workload and reputation incentives are highly relevant when directors contemplate relinquishing an existing directorship.

5 Tests of H2 – Director-Level Analyses: Director Portfolio Adjustments and Reputation Growth

5.1 Identification strategy

To identify the impact of adjustments to a director's portfolio composition (i.e., gaining at least one additional directorship), we use generalized difference-in-difference regressions (DiD) with staggered treatments (similar to Bertrand and Mullainathan 2003; Dube and

[5] Masulis and Mobbs (2014) and Ormazabal (2018) shift their dependent variable (a dummy variable indicating when a director left a firm in a given year) one year forward. Since we shift the reported start and end dates of a directorship by up to +/- 6 months, most observations already include time-lagged control variables. For example, if a directorship starts on the 1st of April 2015, we set the start date to the 31st of December 2014. Consequently, we use control variables from 2014 for start dates ranging from January to June 2015. In case of starting dates ranging from July to December, we set the starting date to the 31st of December of that year and use control variables from that year. Nevertheless, additionally lagging all controls by one year does not change our main results.

Zhu 2021; Guo et al. 2022). The DiD unit and time fixed effects replace the post and treatment-effect variables of a traditional DiD model (Goodman-Bacon 2021). The DiD estimator equals a weighted average of all possible standard DiD estimates between a treated and a control group.

To gain an additional board seat, an outside director needs to be considered and elected by a firm's nomination committee (Duchin et al. 2010). Concerning the treatment, our sample covers (1) directors who never gain or lose any directorships (control group), and (2) directors who accept at least one additional board seat nomination (with the period after the gain event termed *PostGain*). Directors, who make at least one adjustment to their portfolio, make their first additional gain decision at different times. Consequently, our treatment (directors being exposed to and accepting an additional board nomination) is staggered. The control group includes directors, who never make any adjustments to their portfolio (never-treated), and directors, who have not yet made any adjustments but are about to gain a directorship in the future (later-treated). We use the following generalized DiD model to test whether gaining at least one additional directorship impacts future reputation growth:

$$ReputationGrowth_{i,t} = \alpha + \beta_1 PostGain_{i,t} + \sum \beta_k Controls_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t} \qquad (3)$$

where i indexes the director and t indexes the year. The dependent variable R*eputationGrowth* captures the relative change in the director's average portfolio reputation from the current to the prior year. The director's average portfolio reputation is calculated as the average of the total market capitalization of all the firms the outside director holds an active seat in a given year. *PostGain* is an indicator variable equal to 1 starting in the calendar year in which a director makes the first additional gain of a directorship. Compliant with the generalized DiD approach, γ_i represent director fixed effects and δ_t represent year fixed effects.

Controls are a vector of director-specific controls. We use controls established in the literature to explain variations in the market capitalization of firms (Kajüter et al. 2019; Fauver et al. 2017; Akbas et al. 2017). Since we use a director-year portfolio perspective, we calculate the yearly mean for all active firms in the director portfolio. Concerning portfolio-related control variables, we control for firms' market cap, performance (ROA, total asset growth, and Tobin's q), leverage (debt to capital), ownership structure (proportion of closely held shares to total shares), and whether firms are navigating through any issues (approximated by earnings restatements). Furthermore, we control for the independence level of a firm's board and its board size (Coles et al. 2008). Lastly, we lag portfolio controls by one year since their impact on relative *ReputationGrowth* most likely does not materialize immediately.[6] Next, we implement director controls. We include the director's age (applying the natural logarithm) and the director's average tenure of all her board seats in a given year, since they influence director experience but could also reduce productivity (Fedaseyeu et al. 2018). Lastly, we control for the total amount of directorships held and a dummy whether she was active on any committee of all her portfolio firms.

5.2 Descriptive statistics of the director-portfolio perspective

Table 2 displays the descriptive statistics for our director-portfolio perspective (H2). The mean for our dependent variable *ReputationGrowth* is 0.26, indicating a yearly increase of 26% in directors' relative market capitalization over our sample period. Our variable of interest (*PostGain*) has a mean of 0.62. This indicates that we have slightly more observations of (treated) directors, who at least joined one additional board than directors, who have never or not yet joined an additional board (control group). On average, over 90% (0.93) of our directors serve on at least one committee in at least one of their directorships (*is_comittee_member).* On average, 10.6 directors are serving on boards (*board_size_avg*).

[6] Using portfolio controls without lagging them by one year does not change the results (untabulated).

5.3 Difference-in-difference estimates of H2

Table 4 shows the results of our generalized DiD models, including the impact of *PostGain* on *ReputationGrowth.* For H2, we predict that directors, who gain at least one additional outside directorship, can achieve higher growth in their reputation than directors who only hold one outside directorship. Thus, the impact of *PostGain* on *ReputationGrowth* (β_1) should be positive and significant.

[Insert Table 4 about here]

The coefficient of *PostGain* (Model 1) is significantly different from zero (t = 7.25, p < 0.01), indicating that outside directors, who gain at least one additional directorship position, can achieve higher reputation growth than directors who have not (yet) gained an additional directorship. Consequently, the DiD estimate supports H2.

5.4 Additional analyses and robustness checks of H2

Next, we test for treatment intensity by replacing our binary treatment variable with a continuous treatment variable (similar to Acemoglu et al. 2004; Dube and Zhu 2021). Instead of setting the treatment to one in all director-years after a director got and accepted a nomination for the first time, we replace the value with the maximum number of nominations a director accepts over all her observation periods (e.g., if a director accepted four board nominations, we set the value to four instead of one). Results of Table 4 Model 2 show that reputation growth increases with the number of nominations a director receives and accepts (t = 7.30, p < 0.01).

To reduce endogeneity concerns, we re-run our model with a randomly generated pseudo-event replacing our treatment variable (see Cornaggia et al. (2015) for a similar approach). In contrast to our *PostGain* treatment variable, the pseudo treatment should display an insignificant effect on reputation growth. To generate the pseudo-event dummy, we first restrict our sample to not-yet and never-treated observations to remove any impact of the actual treated observations on the pseudo-treatment estimates. Removing treated observations reduces the sample from 30,981 to 11,902 observations. Next, we calculate the ratio of treated to not-yet and never-treated observations of the original sample, with 0.62 of the observations being treated. We use this ratio and randomly set director-year observations to one (e.g., if a join event is randomly selected for a director in 2014, the director-years 2014 to 2019 are set to one). This also results in a ratio of pseudo-treated events to non-events of 0.62. Thereby, we retain the original treated ratio (similar to Cornaggia et al. 2015). Next, we re-run Model 1 with the pseudo-event variable. Table 4 Model 3 shows an insignificant effect of the pseudo-event on reputation growth ($t = 0.39$, $p = 0.698$), reducing potential endogeneity concerns of our DiD approach being caused by spurious trends.

Additionally, we apply propensity score matching to ensure our results are not driven by observable differences between treated and non-treated (control) observations. First, we apply a probit regression to estimate the probability of receiving an additional board seat nomination for the treatment and control groups using director-portfolio characteristics (a*ge_ln, tenure_avg, independence_level_avg, board_size_avg, mcap_avg*). Results of Table 6 show that the means of the director-portfolio characteristics significantly differ between the treated and control firms before applying propensity score matching (all p-values < 0.01). Second, we match each treated observation with one control observation by applying one-to-one nearest neighbor matching without replacement. We use a caliper of 0.4% (equaling twice the standard deviation of the logit of the propensity score, see Dube and Zhu (2021)) to limit the range of potential propensity score matches, resulting in 11,409 matched director-year pairs.

Table 6 shows that the differences in means of the treated and propensity score matched control group become insignificant ($p > 0.1$), except for the lagged portfolio market cap variable ($p = 0.052$). Last, we apply the frequency weights (equaling one for each observation since we use one-to-one matching) of our matched observations to our generalized DiD model (see equation 3), limiting the model to only treated observations with a nearest neighbor match. This results in 22,819 observations (Table 5 Model 1). The coefficient of PostGain still significantly differs from zero ($t = 7.06$, $p < 0.01$), showing that the results are not driven by observable differences in characteristics between the treatment and control groups.

[Insert Table 5 and Table 6 about here]

As an additional analysis, we analyze whether outside directors, who never relinquish one of their multiple board seats, achieve higher reputation growth than those who relinquish at least one of their multiple directorships during our 20-year observation period. Director exits could increase their average reputation by leaving poorly performing firms (Arthaud-Day et al. 2006; Cowen and Marcel 2011) and by allowing them to allocate more time to their remaining monitoring duties (Bar-Hava et al. 2020). However, leaving a firm's board potentially leads to the loss of information and knowledge synergies. Furthermore, since not all exits are voluntary (Boivie et al. 2012), forcefully exiting a well-performing firm would lower a director's reputation. We replace the *PostGain* treatment variable with a dummy (*GainVsRelinquish*), which equals zero for directors who relinquish at least one directorship and one for directors who hold multiple seats and never relinquish one of their directorships.

Results of Table 5 Model 2 show that directors, who never relinquish one of their board seats, achieve higher reputation growth afterward than directors who leave one of their multiple seats ($t = 2.69$, $p = 0.021$).[7] These results substantiate our theory that reputation growth could be mainly driven by information and expertise synergies gained from holding additional directorships, which are lost when leaving a board.

Lastly, two potential issues arise since our BoardEx observations (starting in 1999) potentially include directors who could have held multiple seats in and before 1999. First, some

---

[7] Table 5 Model 2 is subject to a reduced sample size. This is due to the relatively small number of directors, who hold multiple directorships while never relinquishing one of their seats during our 20-year observation period. However, a sample with more than 1,500 observations should still be sufficient to make causal inferences with adequate statistical power. Moreover, it should be noted that we omit director fixed effects in this model since the *GainVsRelinquish* variable is static within director observations and, thus, would be absorbed by the fixed effects.

directors already held multiple seats in 1999 and are thus already treated. Since this could add bias to the treatment effect (Baker et al. 2022; Goodman-Bacon 2021), we remove these directors as a robustness check. This sample adjustment (n = 30,259) yields a similar result of the treatment effect as before ($t = 7.25$, $p < 0.01$, untabulated). Second, our sample might include directors who only held one directorship in 1999 but have gained and relinquished an additional seat before 1999. As an even more conservative robustness check, we remove all directors already active in 1999. Even though the sample size decreases from 30,259 to 21,045, the significance level of the nomination treatment on reputation growth remains unchanged ($t = 5.42$, $p < 0.01$, untabulated).

6 Tests of H3 – Firm-Level Analyses: Board of Directors Incentives and Firm Outcomes

6.1 Identification strategy

To examine the effect of the average reputation (risk propensity) of all outside directors in a given firm on performance (earnings management), we estimate two-way fixed effects regression models:

$$FirmDV_{i,t} = \alpha + \beta_1 BoardIncentive\_avg_{i,t} + \sum \beta_k Controls_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,t} \qquad (4)$$

where i indexes the firm and t indexes the year. For H3a, we use ROA (net income divided by total assets) as the dependent variable and *BoardReputation_avg* as our main independent variable, which captures the average portfolio reputation (market capitalization) of all outside directors in a given firm (i.e., the average of all directors' average portfolio reputation). For H3b, we use earnings management as the dependent variable. We calculate earnings management based on the modified Jones Model (Dechow et al. 1995), where the firm-specific discretionary accruals are estimated from the total accruals. For H2b, our main independent variable is *BoardRisk_avg,* which captures the average portfolio risk (beta) of all outside directors (i.e., the average of all directors' average portfolio risk propensity).

Again, we apply two-way fixed effects regression models, with firm-level fixed effects controlling for time-invariant firm characteristics. We include industry-year-level fixed effects on the second level to control for industry characteristics that could potentially determine firm outcomes. Standard errors are robust to heteroskedasticity and serial correlations within industry-year-level clusters. Lastly, increased performance could lead to firms acquiring more reputable directors. To alleviate potential endogeneity problems arising from reverse causality, we follow prior research (e.g., Barnett and Salomon 2012) by including the lagged dependent variable as an additional independent variable in our regression models.

6.2 Control variables

We use several board-level and firm-level control variables that could influence firm performance or earnings management. We use the share of outside directors on a board that serves on audit committees in any of their directorships (*BoardAuditCom_Share*) as a proxy for audit and financial expertise. Outside directors with the necessary experience to be appointed to audit committees might be more effective in detecting and preventing accrual-based earnings management. Similarly, we control for the share of directors with higher education (*BoardHighEducation_Share*) and the average network size (*BoardNetworkSize_avg*) to proxy for the ability and resources of the outside directors of a given firm. Additionally, we control for the average level of governance in the directorship portfolios of all independent board members (*BoardGovernance_avg*). On the firm level, we control for the change in other accruals as those might be associated with the magnitude of earnings management. We control for the number of outside directors that serve on the board (*Director_Count*) as the independent board size might influence the monitoring effectiveness of its outside directors. Additionally, we include several other firm-level controls, including shareholder structure (*Closelyheldshares*). Table 7 shows the descriptive statistics of the firm-year panel.

[Insert Table 7 about here]

6.3 Results of H3a and H3b

Table 8 Model 1 shows a significant positive coefficient for *BoardReputation_avg* (t = 3.97, p < 0.01), supporting H3a. Firms monitored by more reputable outside directors (based on the average reputation of their directorship portfolios) tend to exhibit higher firm performance measured by return on assets. A potential explanation for this finding is that more reputable outside directors are more incentivized to protect their (higher) reputational capital. Thus, they are more inclined to monitor and advise their directorships effectively.

[Insert Table 8 about here]

Based on H3b, we expect the average risk propensity of a firm's outside directors to influence *EarningsManagement*. We find evidence supporting H3b, as the coefficient for our measure that captures the average directors' portfolio risk (*BoardRisk_avg*) is significant and positive (t = 1.93, p = 0.054). We argue that firms whose outside directors tolerate more risk in their directorship portfolios are also more inclined to accept potentially income-increasing earnings management and bear the increased litigation risk.

6.4 Robustness checks of H3a and H3b

In Table 8 Model 1 (2), we use the average of all directors' average portfolio reputation (risk) in a firm as our proxy for a board's average reputation (risk propensity). However, this absolute measure does not consider the relative ranking of a specific board seat compared to the other firms in a director's portfolio. Therefore, we replace the absolute average measurements with a relative average measure using the average of the *ReputationDummy (RiskDummy*) of all outside directors in a firm, which we already used in H1. We name the variable *ReputationDummy_avg (RiskDummy_avg)*, with a value of one (zero) indicating that all outside directors rank the respective firm higher (lower) than the average of all her other portfolio firms. The results of Model 3 show that when firms employ directors who rank the firm relatively higher in reputation than the average of their director portfolio, firm performance is increased (t = 3.52, $p < 0.01$). A potential explanation for this finding is that if the respective firm has a higher reputation than its directors' other board seats, its directors are incentivized to better monitor and protect the higher-ranked directorship. Thus, the absolute and relative reputations of a firm's outside directors positively influence firm performance.

On the contrary, even though the absolute risk propensity increases the likelihood of earnings management, the coefficient of the *RiskDummy_avg* in Model 4 is insignificant (t = -0.14, p = 0.70). Thus, earnings management is not influenced by the proportion of outside directors for whom the respective firm exhibits more risk than the average of their directorship portfolio. We argue that directors do not change their monitoring behavior of a directorship that has relatively larger risk than the average of his portfolio. Instead, only the absolute risk propensity level of all outside directors influences earnings management.

In conclusion, we find evidence supporting H3a and H3b as our results corroborate the notion that the portfolio reputation and (absolute) risk propensity of all outside directors serving on a firm's board influence firm performance and earnings management.

7 Conclusion

Due to the changes of the Sarbanes-Oxley Act (SOX) in 2002 to the director labor market, researchers started to examine outside directors' incentives (e.g., reputation, risk, and workload-related concerns) (e.g., Boivie et al. 2012; Masulis and Mobbs 2014; Ormazabal 2018). Our research adds to prior literature by examining the interplay between director portfolio adjustments (e.g., relinquishing or joining a board), the underlying incentives, and their impact on firm outcomes.

First, we find that directors strategically relinquish board seats to increase their reputation since they tend to exit their lowest reputable (and most work intense) directorships (H1).

Our results have practical relevance for firms wanting to retain directors by managing their workload and considering the director's relative portfolio reputation when offering a seat.

Second, since prior research mostly focuses on relinquishing decisions (e.g., Boivie et al. 2012; Ormazabal 2018; Masulis and Mobbs 2014), we change the perspective and examine the impact of accepting an additional directorship on directors' reputation growth. Our results show that accepting an additional board seat is advantageous for a director's reputation growth compared to only holding one board seat (H2). Additional analyses reveal that the increased reputation growth could be driven by the information and knowledge-sharing synergies since leaving an additional board seat leads to lower reputation growth afterward. These findings have practical implications for directors contemplating accepting an additional nomination, since accepting is generally advantageous due to future increases in reputational capital.

Lastly, by taking on a firm perspective, we examine the average portfolio characteristics of firms' outside directors and their impact on firm outcomes. We find that a higher average portfolio reputation of independent board members is associated with better firm performance (H3a). Furthermore, our results shed light on prior mixed findings concerning the relationship between the proportion of outside directors and firms' earnings management (e.g., Wang et al. 2015; Klein 2002), since we show that the board's portfolio risk increases earnings management (H3b). Our findings regarding the significant effect of director portfolio attributes on firm outcomes give firms additional insights when making nominations of new outside directors.

There are also several conceivable avenues for further empirical studies. For example, future research could examine whether outside director incentives have any influence on the respective internal directorship of the director. E.g., do increases in outsider reputation lead to career advancements or better performance of the internal directorship? Lastly, it could be of interest to examine whether firms with risk-seeking directorship portfolios of their board of directors lead to adverse circumstances (e.g., financial fraud).

Tables

Table 1 – Descriptive statistics: Director-firm-level analysis

Variable	Mean	Std. Dev.	Min	Max
Relinquished	.065	.247	0	1
ReputationDummy	.552	.497	0	1
RiskDummy	.255	.436	0	1
CompensationDummy	.424	.494	0	1
MeetingsDummy	.416	.493	0	1
GainDummy	.255	.436	0	1
Age	61.910	6.873	31	94
Directorships_Count	2.756	.987	2	14
Tenure	5.67	3.991	0	20
Committee_Nomination	.414	.493	0	1
Committee_Comp	.481	.5	0	1
Committee_Audit	.547	.498	0	1
Committee_Finance	.169	.375	0	1
Committee_Governance	.467	.499	0	1
Committee_Risk	.079	.269	0	1
Committee_Chairman	.555	.497	0	1
Size ($ mio.)	20.05	34.10	6.36	1650
ROA	4.153	13.199	-119.38	34.61
Board_Attendance	81.873	10.064	0	100
Debt	43.252	51.242	-4304.07	1669.37
Closely_Held_Shares	9.715	16.678	0	99.15
Growth	10.354	34.147	-52.2	451.46
TobinsQ	2.084	1.53	.467	15.667

The descriptive statistics are based on a sample of 32,970 observations for the period of 1999 to 2019. A detailed description of all used variables can be found in the appendix.

Table 2

Director incentives and relinquish decisions

Variables	**Model 1** Dummy incentives Relinquished	**Model 2** Relative incentives Relinquished	**Model 3** Dummy incentives + time-lagged controls Relinquished	**Model 4** Relative incentives + time-lagged controls Relinquished
ReputationDummy	-0.014***		-0.017***	
	(-3.120)		(-3.770)	
RiskDummy	0.000		-0.001	
	(0.110)		(-0.250)	
Compensation	-0.004		-0.006	
Dummy	(-1.210)		(-1.610)	
MeetingsDummy	0.013***		0.011***	
	(3.540)		(2.760)	
RelativeReputation		-0.000***		0.000***
		(-2.690)		(-3.489)
RelativeRisk		-0.004		-0.004
		(-1.300)		(-1.41)
Relative		-0.002		-0.003
Compensation		(-0.890)		(-1.547)
RelativeMeetings		0.014***		0.011***
		(4.680)		(3.866)
GainDummy	-0.009**	-0.005	-0.008**	-0.005
	(-2.330)	(-1.310)	(-2.090)	(-1.157)
Age_ln	-2.366***	-2.602***	-3.045***	-3.342***
	(-6.320)	(-6.210)	(-7.570)	(-7.209)
Directorships_	0.019***	0.019***	0.021***	.021***
Count	(6.160)	(5.790)	(6.320)	(6.267)
Tenure	0.008***	0.007***	0.007***	.007***
	(11.820)	(11.000)	(10.920)	(10.286)
Committee_	0.009	0.011*	0.011*	.011*
Nomination	(-1.370)	(1.770)	(1.720)	(1.715)
Committee_Comp	-0.026***	-0.022***	-0.025***	-0.023***
	(-5.300)	(-4.690)	(-5.040)	(-4.536)
Committee_Audit	-0.011**	-0.013***	-0.008	-0.011**
	(-2.250)	(-2.850)	(-1.550)	(-2.175)
Committee_	-0.003	-0.003	-0.006	-0.006
Finance	(-0.450)	(-0.390)	(-0.810)	(-0.813)
Committee_	-0.030***	-0.034***	-0.034***	-0.036***
Governance	(-4.820)	(-5.710)	(-5.177)	(-5.645)
Committee_Risk	-0.014	-0.010	-0.013	-0.014
	(-1.580)	(-1.120)	(-1.351)	(-1.418)
Committee_	-0.031***	-0.030***	-0.031***	-0.029***
Chairman	(-6.540)	(-6.270)	(-6.21)	(-5.912)
Size	-0.004	-0.008***	-0.002	-0.006**
	(-1.430)	(-3.210)	(-0.801)	(-2.439)
ROA	-0.001***	-0.000*	0.000**	0.000**
	(-2.960)	(-1.690)	(-2.203)	(-2.157)
Board_Attendance	-0.000	0.000	0.000	0.000
	(-1.080)	(1.080)	(.588)	(.81)
Debt	-0.000	-0.000	0.000	0.000
	(-1.270)	(-1.070)	(.843)	(1.081)
Closely_Held_	0.000	0.000	0.000	0.000
Shares	(0.950)	(1.510)	(-1.307)	(-0.605)
Growth	-0.000	-0.000	0.000*	0.000
	(-0.170)	(-0.290)	(-1.811)	(-1.233)
TobinsQ	-0.001	-0.001	-0.003	-0.002

	(-0.850)	(-1.050)	(-1.579)	(-1.386)
Director FE	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes
Observations	32,970	29,617	29,742	26,876
Adjusted R-squared	0.0869	0.0935	0.0905	0.0962

This table presents the effect of directorship reputation on the relinquish decision. The dependent variable in both columns (*Relinquished*) is a dummy variable that is set to one if the outside director leaves the firm's board in a given year and zero if she has an active position in the firm's board of directors. The main independent variable *ReputationDummy* in column (1) is a dummy variable that is set to one if the market capitalization of the relinquished firm is larger than the average market capitalization of the remaining firms in the director portfolio and zero otherwise. The main independent variable *RelativeReputation* in Model 2 is a continuous variable that measures the relative difference from the relinquished directorship's market capitalization to the average market capitalization of the remaining firms in the director portfolio. Model 3 (4) is equal to Model 1 (2) except that all control variables are lagged by one year. The continuous independent variables and all financial variables are winsorized at the 1st and 99th percentile. We include director and industry-year fixed effects in all models. We cluster standard errors at the industry-year level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 3 – Descriptive statistics: Director-level analysis

Variable	Mean	Std. Dev.	Min	Max
ReputationGrowth	.263	.989	-0.895	7.013
PostGain	.62	.485	0	1
Age_ln	4.095	.139	3.401	4.564
Tenure_avg	5.432	3.515	0	20
Directorships_count	2.013	1.143	1	24
Is_committee_member	.93	.256	0	1
Size_avg [$ mio.]	15.589	1.586	8.921	193.040
ROA_avg	4.118	11.034	-91.04	29.64
Growth_avg	12.52	33.405	-45.18	309.212
TobinsQ_avg	1.986	1.159	.624	9.864
Debt_to_capital_avg	38.905	311.138	-48898.879	5741.113
Earnings_Restatement_avg	.035	.185	0	1
Closely_held_shares_avg	13.787	16.326	0	99.8
Independence_level_avg	77.542	15.458	0	100
Board_size_avg	10.661	2.368	1	35

The descriptive statistics are based on a sample of 31,283 observations for the period of 1999 to 2019. A detailed description of all used variables can be found in the appendix.

Table 4

Director portfolio adjustments and reputation growth (1)

	Model 1 Generalized DiD	**Model 2** Treatment intensity	**Model 3** Pseudo treatment
Variables	ReputationGrowth	ReputationGrowth	ReputationGrowth
PostGain	0.311***	0.092***	0.009
	(7.250)	(7.304)	(.388)
Age_ln	0.454	0.893	4.322**
	(0.610)	(.699)	(2.432)
Tenure_avg	-0.022***	-0.019***	-0.032***
	(-4.220)	(-3.281)	(-2.715)
Directorships_count	0.010	0.002	-0.161***
	(0.720)	(.152)	(-4.593)
Is_committee_member	0.190***	0.204***	0.159
	(4.540)	(2.93)	(1.217)
Size_avg	-0.641***	-0.633***	-0.504***
	(-14.750)	(-31.792)	(-13.435)
ROA_avg	-0.003*	-0.003**	-0.006***
	(-1.990)	(-2.455)	(-3.359)
Growth_avg	-0.000	0.000	-0.001**
	(-1.350)	(-0.975)	(-2.336)
TobinsQ_avg	0.036**	0.032***	-0.004
	(2.540)	(2.734)	(-0.234)
Debt_to_capital_avg	0.000	0.000***	0.000**
	(1.030)	(3.877)	(2.243)
Earnings_Restatement_avg	-0.014	-0.009	0.003
	(-0.460)	(-0.379)	(.094)
Closely_held_shares_avg	-0.000	0.000	0.002*
	(-0.250)	(-0.058)	(1.948)
Independence_level_avg	0.000	0.000	0.002**
	(0.190)	(.396)	(2.058)
Board_size_avg	0.065***	0.062***	0.027***
	(11.790)	(10.186)	(2.988)
Director FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	31,283	30,981	11,902
Adjusted R-squared	0.2935	0.2892	0.3204

This table presents the effect of directors' portfolio adjustments on director reputation growth. The dependent variable measures the relative change from directors' average market capitalization from the prior to the current year. The treatment variable *PostGain* is set to one as soon as a director gains one additional directorship and is set to zero for never and not yet treated director-years (Model 1). In Model 2 we test for treatment intensity by replacing our binary treatment variable with a continuous treatment variable. In Model 3 we re-run Model 1 with a randomly generated pseudo-event replacing our treatment variable. The dependent variables and all financial variables are winsorized at the 1st and 99th percentile. Firm-level variables are lagged by one year. In all generalized difference-in-difference regressions, we include director and year fixed effects. We cluster standard errors at the director-level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 5

Director portfolio adjustments and reputation growth (2)

Variables	Model 4 Propensity Score Matching ReputationGrowth	Model 5 Gain vs. Relinquish ReputationGrowth
PostGain	0.314***	-
	(7.058)	
GainVsRelinquish	-	0.167**
		(2.690)
Age_ln	1.905	-0.005
	(1.342)	(-0.050)
Tenure_avg	-0.025***	-0.008*
	(-3.572)	(-1.960)
Directorships_count	-0.003	0.024
	(-0.171)	(1.340)
Is_committee_member	.219***	0.103**
	(2.772)	(2.550)
Size_avg	-0.667***	-0.135***
	(-27.775)	(-4.090)
ROA_avg	-0.002*	0.002
	(-1.876)	(0.870)
Growth_avg	-0.000	0.001
	(-0.336)	(0.640)
TobinsQ_avg	0.037***	0.001
	(2.66)	(0.030)
Debt_to_capital_avg	0.000***	0.000
	(4.076)	(0.020)
Earnings_Restatement_avg	-0.016	0.029
	(-0.500)	(0.410)
Closely_held_shares_avg	-0.000	-0.000
	(-0.099)	(-0.090)
Independence_level_avg	0.000	-0.002**
	(0.271)	(-2.180)
Board_size_avg	0.062***	0.018
	(8.578)	(1.080)
Director FE	Yes	No
Year FE	Yes	Yes
Observations	22,819	1,529
Adjusted R-squared	0.2967	0.1258

This table presents the effect of directors' portfolio adjustments on director reputation growth. The dependent variable measures the relative change from directors' average market capitalization from the prior to the current year. For Model 4 we extract frequency weights from propensity score matching and apply them to our generalized Difference-in-Difference model (Model 1), limiting the model to only treated observations with a nearest neighbor match. In Model 5, the *PostGain* treatment variable is replaced with a dummy (*GainVsRelinquish*), which equals zero for directors who relinquish at least one directorship and one for directors, who hold multiple seats and never relinquish one of their directorships. The dependent variables and all financial variables are winsorized at the 1st and 99th percentile. Firm-level variables are lagged by one year. In all generalized difference-in-difference regressions, we include director and year fixed effects. We cluster standard errors at the director-level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 6

Propensity Score Matching Results

Variable (*mean*)	Before Matching			After Matching		
	Treated (n = 18,155)	Unmatched Control (n = 12,104)	t-test	Treated (n =11,409)	Matched Control (n = 11,409)	t-test
Age_ln	4.11	4.07	27.84 ***	4.10	4.10	-1.00
Tenure_avg	5.63	5.15	11.73 ***	5.57	5.71	-0.41
Independence_level_avg	78.05	76.89	6.44 ***	77.13	77.24	-0.56
Board_size_avg	10.71	10.58	4.80 ***	10.59	10.60	-0.51
Size_avg	15.71	15.41	16.48 ***	15.51	15.53	-1.95*

This table presents the means of the treated (*PostGain* equals one) and control groups. Matching is based on one-to-one nearest neighbor matching without replacement using a probit regression to estimate the probability of receiving an additional board seat nomination for the treatment and control groups using director-portfolio characteristics. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 7: Descriptive statistics: Firm-level analysis

Variable	Mean	Std. Dev.	Min	Max
ROA	3.172	17.132	-463.22	289.2
EarningsManagement	.124	1.009	-24.197	15.27
BoardReputation_avg [$ mio.]	20.22	51.02	5	896.70
BoardRisk_avg	1.246	.728	-2.824	21.505
BoardAuditCom_Share	.544	.362	0	1
BoardHighEducation_Share	.586	.356	0	1
BoardNetworkSize_avg	595.027	1175.256	0	17528
BoardExec_Share	.152	.275	0	1
BoardGovernance_avg	23.521	26.145	0	96.75
BoardDebt_avg	27.955	322.846	-19522.078	2113.33
LnAge_avg	4.056	.137	3.332	4.511
NetIncomeGrowth	58.138	1956.761	-99.93	182891.67
ChangeOtherAccruals	10560.03	160924.9	-2091000	7781000
Director_Count	2.879	1.81	1	14
Closelyheldshares	18.981	20.378	0	100
Amort_Intangibles	42483.409	185121.08	-17262	7231000
TotalAssets	5147943.8	16772485	535	4.072e+08

The descriptive statistics are based on a sample of 8,289 observations for the period of 1999 to 2019. A detailed description of all used variables can be found in the appendix.

Table 8

Board of director incentives and firm outcomes

Variables	**Model 1** Absolute Board Reputation RoA	**Model 2** Absolute Board Risk Earnings Management	**Model 3** Relative Board Reputation RoA	**Model 4** Relative Board Risk Earnings Management
BoardReputation_avg	0.000***	-	2.546***	-
	(3.966)		(3.58)	
BoardRisk_avg	-	0.033*	-	-.014
		(1.929)		(-.385)
BoardAuditCom_	1.434**	0.029	0.854	.055
Share	(1.910)	(0.910)	(0.941)	(1.038)
BoardHighEducation_	-0.046	-0.019	0.400	.008
Share	(-0.050)	(-0.390)	(0.336)	(.132)
BoardNetworkSize_	-0.000	-0.000	-0.000**	-0.000
avg	(-1.320)	(-1.540)	(-2.276)	(-1.204)
BoardExec_Share	-1.262	-0.027	-1.428	-.02
	(-1.240)	(-0.780)	(-1.312)	(-.423)
BoardGovernance_avg	-0.003	-0.000	-0.006	0.000
	(-0.260)	(-0.590)	(-0.567)	(.288)
BoardDebt_avg	0.000	0.000	0.000	0.000
	(0.310)	(0.320)	(.741)	(.097)
LnAge_avg	-7.399***	0.209*	-2.65	.47**
	(-3.110)	(1.730)	(-0.986)	(2.341)
NetIncomeGrowth	-0.000	-0.000	-0.000***	-0.000**
	(-1.620)	(-0.950)	(-3.631)	(-2.28)
ChangeOtherAccruals	0.000**	-0.000	0.000	0.000
	(1.960)	(-0.770)	(1.201)	(.402)
Director_Count	-2. 254**	-0.007	-0.332**	-.002
	(-1.820)	(-0.950)	(-2.384)	(-.2)
Closelyheldshares	-0.063***	0.000	-0.036	.002*
	(-3.150)	(0.660)	(-1.472)	(1.652)
Amort_Intangibles	-0.000***	0.000	-0.000***	0.000
	(-2.850)	(0.820)	(-3.14)	(1.364)
TotalAssets	-0.000***	-0.000	-0.000**	-0.000*
	(-3.280)	(-1.080)	(-2.563)	(-1.864)
ROA_lagged	0.622		0.119***	
	(1.550)		(3.028)	
EarningsManagement_		0.002*		-.106
lagged		(1.760)		(-1.433)
Firm FE	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes
Observations	8,289	8,395	5,928	5,934
Adjusted R-squared	0.4078	0.3695	0.4962	0.5125

This table presents the effect of the board of directors' portfolio reputation and risk on firm outcomes. The dependent variable in Model 1 and 3 is the return on assets (RoA). The dependent variable in Models 2 and 4 is accrual-based earnings management based on the Modified Jones Model (Dechow et al., 1995). The main independent variables of Models 1 and 2 (*BoardReputation_avg* and *BoardRisk_avg)* capture the average portfolio reputation (risk) of all outside directors in a given firm (i.e., the average of all directors' average portfolio reputation). In Models 3 and 4 the absolute average measurements are replaced with a relative average measure using the average of the *ReputationDummy* (*RiskDummy*) of all outside directors in a firm, with values of one (zero) indicating that all outside directors rank the respective firm higher (lower) than the average of all her other portfolio firms. We include firm and industry-year fixed effects. We cluster standard errors at the industry-year level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Appendix A

Variable Definitions

Variable	Definition
Director Portfolio Changes	
Relinquished	Indicator variable that equals one if the outside director leaves the firm's board of directors in a given year, and zero if she has an active position in the firm's board of directors
PostGain	Indicator variable that is set to one in the calendar year and all subsequent years, in which the director joins at least one additional board, and zero otherwise.
GainVsRelinquish	Indicator variable that is set to one (zero) for directors who hold multiple directorships and never relinquished an existing directorship.
Director Incentives	
Size (market capitalization as reputation proxy)	Common shares outstanding multiplied with the stock price
Risk (beta as risk proxy)	Month-end price percent changes and their relativity to the local market index
Compensation (as compensation proxy)	The average compensation of the board members in US dollars.
Meetings (workload proxy)	The number of board meetings during the year.
H1	
RelativeReputation	The firm's market capitalization in relation to the average total market capitalization of the director's board of director portfolio excluding the firm
RelativeRisk	same as above with beta instead of market cap
RelativeCompensation	same as above with compensation instead of market cap
RelativeMeetings	same as above with meetings instead of market cap
ReputationDummy	Dummy variable that equals one if the firm's market capitalization is larger than the average market capitalization of the director's board of director portfolio excluding the firm
RiskDummy	same as above with beta instead of market cap
CompensationDummy	same as above with compensation instead of market cap
MeetingsDummy	same as above with meetings instead of market cap

H2	
ReputationGrowth	The relative change of the director's portfolio average market capitalization from the previous to the current year
H3	
BoardReputation_avg	Average market capitalization of the directorship portfolios of all outside directors that serve a given firm (average of all directors average portfolio reputation)
BoardRisk_avg	same as above with firm beta instead of market cap
ReputationDummy_avg	Average of the ReputationDummy (equaling 0 or 1) of all outside directors that serve a given firm
RiskDummy_avg	Average of the RiskyDummy (equaling 0 or 1) of all outside directors that serve a given firm
Director Controls	
GainDummy	Dummy variable that equals one if the director gained at least one board of director position during the respective year
Age_ln	Natural logarithm of the director's age in the respective year
Directorships_Count	Amount of outside directorships a director holds in the respective year
Tenure	The director's tenure in the firm measured in years
Committee_Chairman	Indicator variable that equals one if the director was active as the chairman of the firm in a given year and zero otherwise
Committee_Nomination	Indicator variable that equals one if the director was active in the nomination committee of the firm in a given year and zero otherwise
Committee_Comp	Indicator variable that equals one if the director was active in the compensation committee of the firm in a given year and zero otherwise
Committee_Audit	Indicator variable that equals one if the director was active in the audit committee of the firm in a given year and zero otherwise
Committee_Finance	Indicator variable that equals one if the director was active in the finance committee of the firm in a given year and zero otherwise
Committee_Governance	Indicator variable that equals one if the director was active in the governance committee of the firm in a given year and zero otherwise
Committee_Risk	Indicator variable that equals one if the director was active in the risk committee of the firm in a given year and zero otherwise
Firm Controls	
Size	The firm's market capitalization (market cap)
ROA	(Net Income – Bottom Line + ((Interest Expense on Debt-Interest Capitalized) * (1-Tax Rate))) / Average of Last Year's and Current Year's Total Assets * 100
Growth	(Current Year's Total Assets / Last Year's Total Assets - 1) * 100

TobinsQ	(Total Assets – Book Equity + Market Value of Equity) / Total Assets
Board_attendance	The average overall attendance percentage of board meetings as reported by the company
Debt	(Long Term Debt + Short Term Debt & Current Portion of Long Term Debt) / (Total Capital + Short Term Debt & Current Portion of Long Term Debt) * 100
Closely_held_shares	Shares held by insiders
Independence_level	Percentage of independent board members as reported by the company
Board_size	The total number of board members
NetIncomeGrowth	Yearly total growth of net income by a firm.
ChangeOtherAccruals	Yearly total change of other accruals disclosed in cash flow statements
Amort_Intangibles	Total amount of amortization of intangible assets (e.g., patents) by year
EarningsManagement	Estimation of earnings management by discretionary accruals. Computed via the modified Jones Model (Dechow et al. 1995)
Director-Portfolio Controls	
Tenure_avg	The average of the director's tenure of her portfolio directorships measured in years
Is_committee_member	Dummy variable that equals one if the director was active in at least one committee of one of her portfolio firms in a given year
Mcap_avg	Average market cap of all the director's firms in a given year
ROA_avg	Average of the ROA-variable of all the director's firms in a given year
Debt_to_capital_avg	Average of the Debt-variable of all the director's firms in a given year
Earnings_restatement_avg	Average of all director firms of the indicator variable that equals one if the company is in the process of a material earnings restatement in a given year
Growth_avg	Average of the NetIncomeGrowth variable of all the director's firms in a given year
Closely_held_shares_avg	Average of the closely_held_shares-variable of all the director's firms in a given year
TobinsQ_avg	Average of the tobinsQ-variable of all the director's firms in a given year
Independence_level_avg	Average of the independence_level-variable of all the director's firms in a given year
Board_size_avg	Average of the ROA variable of all the director's firms in a given year
Board-level Controls	
BoardAuditCom_Share	Share of all outside directors of a given firm that serve on an audit committee in any of their other directorships

BoardHighEducation_Share	Share of all outside directors of a given firm that possess a higher education (masters or above)
BoardNetworkSize_avg	Average number of network size of all outside directors that serve on the board of a given firm
BoardExec_Share	Share of all outside directors of a given firm that simultaneously serve as an inside (i.e., executive) director on another firm
BoardGovernance_avg	Average governance score of the directorship portfolios of all outside directors that serve a given firm. The governance score is acquired from Refinitv and measures the quality (from 0 to 100) of the corporate governance of a firm.
BoardDebt_avg	Average level of the Debt-variable of the directorship portfolios of all outside directors that serve a given firm.
LnAge_avg	Average age of all outside directors that serve a given firm
Director_Count	Number of independent directors that sit on a board of a given firm

References

Acemoglu, D., Autor, D., & Lyle, D. (2004). Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury. Journal of Political Economy, 112(3), 497–551.

Adams, R. B., & Ferreira, D. (2008). Do directors perform for pay? Journal of Accounting and Economics, 46, 154–171.

Akbas, F., Jiang, C., & Koch, P. D. (2017). The Trend in Firm Profitability and the Cross-Section of Stock Returns. The Accounting Review, 92, 1–32.

Arthaud-Day, M. L., Certo, S. T., Dalton, C. M., & Dalton, D. R. (2006). A Changing of the Guard: Executive and Director Turnover Following Corporate Financial Restatements. Academy of Management Journal, 49, 1119–1136.

Badolato, P. G., Donelson, D. C., & Ege, M. (2014). Audit committee financial expertise and earnings management: The role of status. Journal of Accounting and Economics, 58, 208–230.

Baker, A., Larcker, D. F., & Wang, C. C. Y. (2022). How Much Should We Trust Staggered Difference-In-Differences Estimates? Journal of Financial Economics, 144, 370–395.

Balsmeier, B., Buchwald, A., & Dilger, A. (2015). Executive Turnover and Outside Directors on Two-Tiered Boards. Managerial and Decision Economics, 36, 158–176.

Bar-Hava, K., Gu, F., & Lev, B. (2020). Market Evidence on Investor Preference for Fewer Directorships. Journal of Financial and Quantitative Analysis, 55, 931–954.

Barnett, M. L., & Salomon, R. M. (2012). Does it pay to be really good? addressing the shape of the relationship between social and financial performance. Strategic Management Journal, 33, 1304–1320.

Beneish, M. D. (2001). Earnings management: a perspective. Managerial Finance, 27, 3–17.

Bertrand, M., & Mullainathan, S. (2003). Enjoying the Quiet Life? Corporate Governance and Managerial Preferences. Journal of Political Economy, 111, 1043–1075.

Boivie, S., Graffin, S. D., & Pollock, T. G. (2012). Time for Me to Fly: Predicting Director Exit at Large Firms. Academy of Management Journal, 55, 1334–1359.

Booth, J. R., & Deli, D. N. (1996). Factors affecting the number of outside directorships held by CEOs. Journal of Financial Economics, 40, 81–104.

Boyd, B. (1990). Corporate linkages and organizational environment: A test of the resource dependence model. Strategic Management Journal, 11, 419–430.

Brickley, J. A., Linck, J. S., & Coles, J. (1999). What happens to CEOs after they retire? New evidence on career concerns, horizon problems, and CEO incentives. Journal of Financial Economics, 52(3), 341–377.

Bryan, D. B., & Mason, T. W. (2020). Independent director reputation incentives, accruals quality and audit fees. Journal of Business Finance & Accounting, 47, 982–1011.

Callahan, W. T., Millar, J. A., & Schulman, C. (2003). An analysis of the effect of management participation in director selection on the long-term performance of the firm. Journal of Corporate Finance, 9, 169–181.

Chen, L., & Moers, F. (2018). The market for independent directors. Corporate Governance: An International Review, 26, 429–447.

Coles, J., Daniel, N., & Naveen, L. (2008). Boards: Does one size fit all. Journal of Financial Economics, 87, 329–356.

Cornaggia, J., Mao, Y., Tian, X., & Wolfe, B. (2015). Does banking competition affect innovation? Journal of Financial Economics, 115, 189–209.

Cowen, A. P., & Marcel, J. J. (2011). Damaged Goods: Board Decisions to Dismiss Reputationally Compromised Directors. Academy of Management Journal, 54, 509–527.

Dechow, P. M., Sloan, R. G., & Sweeney, A. P. (1995). Detecting Earnings Management. The Accounting Review, 70(2), 193–225.

Deng, M., Ho, J. L., & Li, S. (2018). Does Managerial Risk Aversion Affect Earnings Management? Evidence from CEO Political Ideology. SSRN Electronic Journal.

Dube, S., & Zhu, C. (2021). The Disciplinary Effect of Social Media: Evidence from Firms' Responses to Glassdoor Reviews. Journal of Accounting Research, 59(5), 1783–1825.

Duchin, R., Matsusaka, J. G., & Ozbas, O. (2010). When are outside directors effective? Journal of Financial Economics, 96, 195–214.

Fahlenbrach, R., Low, A., & Stulz, R. M. (2010). Why do firms appoint CEOs as outside directors? Journal of Financial Economics, 97, 12–32.

Fama, E. F., & Jensen, M. C. (1983). Separation of Ownership and Control. The Journal of Law and Economics, 26, 301–325.

Fauver, L., Hung, M., Li, X., & Taboada, A. G. (2017). Board reforms and firm value: Worldwide evidence. Journal of Financial Economics, 125, 120–142.

Fedaseyeu, V., Linck, J. S., & Wagner, H. F. (2018). Do qualifications matter? New evidence on board functions and director compensation. Journal of Corporate Finance, 48, 816–839.

Ferris, S. P., Jagannathan, M., & Pritchard, A. C. (2003). Too Busy to Mind the Business? Monitoring by Directors with Multiple Board Appointments. The Journal of Finance, 58, 1087–1111.

Fich, E. M. (2005). Are Some Outside Directors Better than Others? Evidence from Director Appointments by Fortune 1000 Firms*. The Journal of Business, 78, 1943–1972.

Fich, E. M., & Shivdasani, A. (2004). Are Busy Boards Effective Monitors? SSRN Electronic Journal.

Fich, E. M., & Shivdasani, A. (2007). Financial fraud, director reputation, and shareholder wealth. Journal of Financial Economics, 86, 306–336.

Fredriksson, A., Kiran, A., & Niemi, L. (2020). Reputation Capital of Directorships and Demand for Audit Quality. European Accounting Review, 29, 901–926.

Ghannam, S., Bugeja, M., Matolcsy, Z. P., & Spiropoulos, H. (2019). Are Qualified and Experienced Outside Directors Willing to Join Fraudulent Firms and If So, Why? The Accounting Review, 94, 205–227.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics, 225, 254–277.

Guo, X., Chen, S., Yu, W., & Liu, C. (2022). Product market competition and controlling shareholders' tunneling: Evidence from China. Managerial and Decision Economics, 43, 3820–3832.

Healy, P. M., & Wahlen, J. M. (1999). A Review of the Earnings Management Literature and Its Implications for Standard Setting. Accounting Horizons, 13, 365–383.

Hossain, M. A., & Oon, E. Y. N. (2022). Board leadership, board meeting frequency and firm performance in two-tier boards. Managerial and Decision Economics, 43, 862–879.

Ibrahim, S., Xu, L., & Rogers, G. (2011). Real and accrual-based earnings management and its legal consequences. Accounting Research Journal, 24, 50–78.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. Journal of Financial Economics, 3, 305–360.

Kajüter, P., Klassmann, F., & Nienhaus, M. (2019). The Effect of Mandatory Quarterly Reporting on Firm Value. The Accounting Review, 94, 251–277.

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

Kor, Y. Y., & Sundaramurthy, C. (2009). Experience-Based Human Capital and Social Capital of Outside Directors. Journal of Management, 35, 981–1006.

Linck, J. S., Netter, J. M., & Yang, T. (2008). The determinants of board structure. Journal of Financial Economics, 87, 308–328.

Linck, J. S., Netter, J. M., & Yang, T. (2009). The Effects and Unintended Consequences of the Sarbanes-Oxley Act on the Supply and Demand for Directors. Review of Financial Studies, 22, 3287–3328.

Lo, K. (2008). Earnings management and earnings quality. Journal of Accounting and Economics, 45, 350–357.

Mace, M. L. (1986). Directors: myth and reality (Harvard Business School classics, Vol. 4). Boston, Mass.: Harvard Business School Pr.

Masulis, R. W. (2020). A Survey of Recent Evidence on Boards of Directors and CEO Incentives. SSRN Electronic Journal.

Masulis, R. W., & Mobbs, S. (2014). Independent director incentives: Where do talented directors spend their limited time and energy? Journal of Financial Economics, 111, 406–429.

Masulis, R. W., Wang, C., & Xie, F. (2012). Globalizing the boardroom—The effects of foreign directors on corporate governance and firm performance. Journal of Accounting and Economics, 53, 527–554.

Mobbs, S. (2013). CEOs Under Fire: The Effects of Competition from Inside Directors on Forced CEO Turnover and CEO Compensation. Journal of Financial and Quantitative Analysis, 48, 669–698.

Ormazabal, G. (2018). Are directors more likely to relinquish their riskiest directorships after the Financial Crisis? Journal of Corporate Finance, 53, 1–20.

Palmrose, Z.-V., Richardson, V. J., & Scholz, S. (2004). Determinants of market reactions to restatement announcements. Journal of Accounting and Economics, 37, 59–89.

Peyer, U. C., & Perry, T. T. (2005). Board Seat Accumulation by Executives: A Shareholder‘s Perspective. Journal of Finance, 60, 2083–2123.

Shiah-Hou, S.-R., & Cheng, C.-W. (2012). Outside director experience, compensation, and performance. Managerial Finance, 38, 914–938.

Shivdasani, A. (1993). Board composition, ownership structure, and hostile takeovers. Journal of Accounting and Economics, 16, 167–198.

Sila, V., Gonzalez, A., & Hagendorff, J. (2017). Independent director reputation incentives and stock price informativeness. Journal of Corporate Finance, 47, 219–235.

Wang, C., Xie, F., & Zhu, M. (2015). Industry Expertise of Independent Directors and Board Monitoring. Journal of Financial and Quantitative Analysis, 50(5), 929–962.

Yermack, D. (2004). Remuneration, Retention, and Reputation Incentives for outside Directors. Journal of Finance, 59(5), 2281–2308.

Do Opposites Attract? The Effect of Cultural Distance on Mergers and Acquisitions: Evidence from Glassdoor Reviews

Authors Andreas Bausch, Marius Brede, Hannes Gerstel, Arnt Wöhrmann

Abstract This paper examines the effects of differences in organizational culture of M&A deals on transaction phase and post-merger outcomes. By applying a state-of-the-art large language model, we generate a novel organizational culture measure based on employee reviews from the Glassdoor website covering M&A deals from 2008 to 2021. Based on the cultural friction hypothesis, we first find that the capital market negatively reacts to M&A announcements of firms with different cultural market orientations. Moreover, our analyses reveal that differences in the emphasis on teamwork and collaboration of acquirers and targets decrease post-merger operating growth. Second, we show that cultural distance is positively related to acquisition premiums, as acquirers with a limited market orientation overestimate the potential synergy realizations of acquiring a target with a strong market orientation. Third, we find that cultural distance does not impact deal durations, except when the acquirer possesses a significantly higher hierarchical orientation than its target. Our findings add to the broad M&A literature and have practical relevance for firms engaging in M&A transactions.

Publication Status Working Paper

Preparing for submission

Acknowledgements We thank Maximilian Schmidt and Sophie Kulig for their helpful comments regarding the conceptualization of the research question, and Valerie Roush for giving us permission to use the Glassdoor data. Lastly, we thank Marco Rausch for his excellence research assistance.

Author Contribution	Bausch	Brede	Gerstel	Wöhrmann
Relative Share	*5%*	*40%*	*50%*	*5%*
Conceptualization	✓	✓	✓	✓
Theory Development		✓	✓	
Methodology		✓	✓	
Data Acquisition		✓	✓	
Data Analyses		✓	✓	
Writing		✓	✓	

1 Introduction

Mergers & Acquisitions (M&A) have become one of the most important strategic initiatives for companies, reaching a total transaction volume of $3.7 trillion in 2017 (Cho and Chung 2022). However, nearly half of all M&A transactions fail (Cartwright and Cooper 1993). In this vein, a survey of senior executives indicates that nearly half would not pursue an M&A target that did not fit their organization's culture, suggesting that cultural differences are among the most common factors in deal failure. This echoes anecdotal evidence attributing cultural friction between acquirers and targets to the failure of high-profile M&A deals such as HP & Compaq and Amazon & Whole Foods (Oberoi 2020).

The management literature generally distinguishes between national and organizational cultures (Rottig 2017). Both types of culture are considered distinct phenomena with their own manifestations and potentially different effects on the organization and its actions (Kirkman et al. 2006). Many studies have empirically examined the effects of national cultural distance on M&A outcomes (e.g., Ahern et al. 2015; Lee 2018; Lim et al. 2016). However, research on the impact of organizational culture on M&A success has been inconclusive due to difficulties in measuring organizational culture (Renneboog and Vansteenkiste 2019; Rottig 2017).

We address this gap in the literature by analyzing the impact of organizational cultural distance between M&A firms on M&A outcomes by directly inferring our cultural distance proxy from employee ratings on Glassdoor. The intelligence website Glassdoor allows employees to anonymously review their employers by providing pros and cons about their employer as textual input. Generally, we expect different impacts of cultural distance on the M&A transaction and post-merger integration phases.

First, we hypothesize that the capital market reacts negatively to announcements of M&A transactions between culturally distant firms (H1a) and that cultural distance leads to lower realized post-merger synergies (H1b). Previous research provides conflicting evidence on the impact of organizational culture differences on M&A success (Rottig 2017; Stahl and Voigt 2008). On the one hand, the cultural learning hypothesis promotes a positive effect on M&A success by suggesting that different cultures exhibit significant learning potential and allow for the profitable recombination of resources (Sørensen 2002). On the other hand, the cultural friction hypothesis (Hofstede 1980) posits that cultural differences between acquirers and targets can lead to additional integration and coordination costs during the (post-) merger phase and thus reduce the success of M&A transactions (Vaara 2002; Weber 1996). We expect induced cultural frictions to outweigh potential learning effects for culturally distant M&A pairs, leading to lower capital market reactions and post-merger synergies.

Second, we hypothesize that cultural differences increase acquisition premiums (H2). Similar to the mixed evidence on cultural distance and M&A announcement returns, prior research also finds mixed evidence on the impact of national culture differences on acquisition premiums (Lim et al. 2016). We argue that cultural differences increase information asymmetries between the acquirer and the target. In addition, they create potential complications in correctly assessing the target's prevailing values and norms. Consequently, we expect that acquirers tend to overestimate the potential synergy realizations of culturally distant targets, leading to overpayments. Third, we hypothesize that cultural differences lead to longer deal durations by increasing the likelihood of protracted negotiations and hold-ups (H3).

Previous studies, such as Rottig's (2017) meta-analysis, attribute these inconclusive findings to the complexity of the culture construct and the inherent methodological weaknesses of using small survey studies as a proxy for firm culture. We overcome these limitations by applying deep-learning-based natural language processing to analyze approximately 400,000 employee reviews of 439 companies (243 M&A deals) on Glassdoor between 2008 and 2021. Our approach is based on the notion that Glassdoor reviews reflect the experiences of individual employees within a company, providing insight into the company's culture (Corritore et al.

2020). Effects of organizational culture are primarily studied using the widely used Competing Values Framework (CVF, Cameron et al. 2006; Quinn and Rohrbaugh 1983). The CVF represents four types of organizational cultural dimensions that are not mutually exclusive: adhocracy, clan, market, and hierarchy (see Appendix A for examples based on Glassdoor reviews). Using state-of-the-art large language models (Bochkay et al. 2023; Vaswani et al. 2017), we apply CultureBERT (Koch and Pasch 2022) to determine our cultural distance score by calculating the Euclidean distance between the four organizational culture dimensions computed from a firm's Glassdoor reviews. In addition to the aggregated cultural distance measure, we calculate four separate measures using the absolute differences in the four culture dimensions. Our analyses provide several insights using the cultural distance score as our main proxy to explain merger outcomes.

First, aggregated organizational cultural distance between acquirers and targets is not negatively associated with capital market reactions (H1a) and post-merger synergies (H1b). However, differences in the market orientation between M&A pairs corroborate the cultural friction hypothesis since higher market differences significantly decrease announcement returns (partially supporting H1a). Moreover, results show that differences in the emphasis on collaboration, teamwork, and employee development (clan orientation) of M&A pairs decrease the acquirer's post-merger operating growth (partially supporting H1b).

Second, our results suggest that cultural differences increase acquisition premiums by negatively affecting the acquirer's ability to determine the target's true value. This effect is mainly driven by acquirers with a limited market culture, who tend to overestimate potential synergy realizations of acquiring a target with a strong market culture. Third, we find no evidence that organizational cultural distance affects the speed with which M&A transactions are completed. However, additional analysis shows that when the acquirer has a significantly higher hierarchy culture score than the target, the deal closes significantly faster. This could be driven by coordination advantages due to reduced opportunities for time-consuming debates caused by the target. Lastly, we run several robustness checks to validate our results.

To the best of our knowledge, our study is the first to apply transformer-based natural language processing to a large sample of Glassdoor reviews to infer the impact of organizational cultural distance on M&A outcomes. In doing so, we contribute to the broader M&A literature in several ways:

First, our study addresses the criticism regarding the predominant use of (small-scale) subjective measures (e.g., self-reported surveys and interviews) to determine organizational culture and its impact on M&A success (Rottig 2017; Teerikangas and Very 2006). Relying on subjective measures increases the risk of common method bias, reduces objectivity, and complicates the comparability of findings. We address this criticism by directly inferring our cultural distance proxy from thousands of voluntarily written employee reviews to provide a more representative understanding of the prevailing organizational culture (Campbell and Shang 2021). Using large textual data, such as annual reports and employee reviews, is an increasingly popular method in the management literature for inferring organizational culture (e.g., Campbell and Shang 2021; Corritore 2018; Li et al. 2021). Using state-of-the-art transformer models, we overcome caveats of prior natural language processing methods (e.g., the inability to reflect the immediate context of surrounding words fully). CultureBERT is up to 28% more accurate at classifying cultural dimensions than previous methods (Koch and Pasch 2022).

Second, we shed light on the inconclusive findings of prior research on the impact of organizational cultural differences on short- and long-term M&A performance (Rottig 2017; Stahl and Voigt 2008). Our results indicate that differences in market and clan orientations impact capital market reactions and post-merger synergies. Regarding capital market reactions, Alexandridis et al. (2022) and Bereskin et al. (2018) find that larger differences in M&A pair's attitudes toward corporate social responsibility (CSR) lead to lower announcement returns. In contrast to our study, the authors use the KLD (EIRIS) CSR score. The score covers a broad range of CSR dimensions, e.g., diversity, environment, human rights, and product. Thus, topics unrelated to corporate culture are also covered in the score. Furthermore, contrary to our cultural distance score, dimensions in the KLD (EIRIS) data do not reflect a continuous probability (intensity) toward the respective dimension, leading to a less granular measurement.

Third, our study contributes to the literature on the importance of acquisition premiums and their impact on the success or failure of M&As (King et al. 2021). To our best knowledge, there is only scarce evidence on the impact of national cultural differences on acquisition premiums (Lim et al. 2016) and no evidence on the effect of organizational cultural distance on premiums. Our results suggest that cultural distance prevents acquiring firms from correctly assessing the true value of the target firm, leading to inflated acquisition premiums. Lastly, we substantiate this finding by revealing that acquirers with a low market orientation overestimate the potential synergy realizations of acquiring a target with a high market orientation.

Furthermore, we are one of the few studies to examine the effect of cultural distance on deal duration. In contrast to Bereskin et al. (2018), who find that high CSR similarity decreases deal duration, our results suggest an insignificant relationship for the overall effect of cultural distance on deal duration. Hence, our results align with Lawrence et al. (2021), who show that national cultural distance does not affect the deal duration. We add to these scarce and mixed findings by showing that differences in hierarchies between the acquirer and target lead to significantly faster deal closing. Thus, instead of spending resources on the uncertainty of deal completion, the acquirer and target can use this time to focus on post-merger integration.

Finally, we add value to M&A practitioners by highlighting the role of cultural differences in M&As by showing that ignoring these cultural differences can have significant (negative) economic consequences, e.g., lower synergy gains and larger acquisition premiums.

Thus, we suggest that acquiring firms should carefully assess differences between their and the target's cultural orientation during and before engaging in M&A transactions.

In conclusion, we extend the managerial literature on the impact of organizational cultural distance on M&A success. The remainder of the paper is organized as follows. In Section 2, we develop our hypotheses. Section 3 presents our data sample. Sections 4 and 5 present the identification strategy and results, and Section 6 provides concluding remarks.

2 Hypotheses Development

2.1 Cultural distance, capital market reactions, and post-merger synergies

The relationship between culture and M&A performance can be decomposed into short- and long-term performance effects. On the one hand, the empirical literature shows that investors consider cultural aspects when evaluating M&A transactions, which is reflected by capital market reactions on the day of the deal announcement (Aktas et al. 2011). On the other hand, transactions lead to changes in profitability, which are reflected in the post-merger integration operating performance (Zollo and Meier 2008). Therefore, we separately derive the effect of cultural distance on capital market reactions (H1a) and post-merger synergies (H1b).

Culture is generally defined as "the collective programming of the mind that distinguishes members of one group or category of people from another" (Hofstede 2001). It is an informal institution that consists primarily of unwritten social rules, values, and norms shaped by its members' shared history and experience (Louis 1981; Schein 1985). Thus, culture significantly influences social interactions by shaping expectations and determining acceptable social behavior (Hofstede 1980).

A central function of organizations is to organize and coordinate the activities of large numbers of people. Consequently, they need control systems to set goals, assess deviations, and provide feedback to individuals. Social control mechanisms, which operate through norms or social expectations, can effectively ensure compliant behavior of organizational members

(Chatman and O’Reilly 2016). Numerous studies show that culture influences organizational practices such as behavioral norms, decision-making, and strategic initiatives such as outsourcing (Dahlgrün and Bausch 2019), CSR (Chen and Liu 2022), innovation (Büschgens et al. 2013) and M&As (Bhagat and McQuaid 1982; Kirkman et al. 2006; Lorsch 1986). However, recent meta-analyses have failed to clarify whether cultural aspects positively or negatively impact acquisition performance (Rottig 2017; Stahl and Voigt 2008).

Arguments for the positive impact of cultural differences on M&A success are primarily based on the theory of interorganizational learning (Sørensen 2002). According to this theory, M&As with culturally distant firms offer significant learning potential and allow for the profitable recombination of resources. Acquiring a company with a different culture facilitates knowledge transfer and access to new practices and techniques (Chakrabarti et al. 2009; Morosini et al. 1998; Sarala and Vaara 2010), creating a sustainable source of value creation (Bouwman 2013; Haspeslagh and Jemison 1991).

Finally, based on the information processing hypothesis, culturally distant acquirers and targets can benefit from multiple perspectives (Kogut and Singh 1988; Watson et al. 1993). This leads to enhanced problem-solving skills, creativity, innovation, adaptability, and ultimately improved performance. Supporting the cultural learning hypothesis, Ahern et al. (2015), Morosini et al. (1998), Conn et al. (2005) and Chakrabarti et al. (2009) find a positive relationship between national cultural distance, announcement returns and post-merger performance.

In contrast to the cultural learning hypothesis, cultural differences between acquirers and targets can be a source of friction during the (post-) merger phase (Vaara 2002; Weber 1996). These frictions can result from personality clashes, incompatible organizational structures and processes, and difficulties transferring core competencies and knowledge between firms (Rottig 2017).

Stahl and Voigt (2008) also draw upon the Social Identity Theory (Tajfel 1981; Turner 1982), which suggests that organizational members are biased toward in-group members. At the same time, they tend to evaluate out-group members negatively to increase the relative status of their own group. This negative impact on the internal cohesion of the workforce can reduce trust between different groups (Sitkin and Stickel 1996), which increases the potential for conflict (Ahern et al. 2015; Jehn et al. 1999; Martin 1992). In addition, internal tensions within the workforce can impede the flow of information between members of the acquirer and the target. As a result, employee and stakeholder resistance can evolve, hindering efficient decision-making (Akerlof 1997; Arrow 1974; Renneboog and Vansteenkiste 2019).

The above factors are associated with higher coordination costs and, ultimately, lower firm performance (e.g., Ahern et al. 2015; Aybar and Ficici 2009; Barkema et al. 1996; Rahahleh and Wei 2013; Weber and Camerer 2003). In this sense, Cartwright and Cooper (1993) show that administrative conflicts can arise from acquiring a culturally distant firm. Moreover, cultural distance during the integration phase can lead to discomfort and hostility among employees of the acquirer and target (Buono et al. 1985).

Several previous studies find evidence supporting the culture friction hypothesis. For example, Chatterjee et al. (1992) find a negative relationship between cultural differences and M&A announcement effects. Similarly, multiple studies find a negative impact on announcement and long-term returns due to national cultural differences between acquirer and target countries (Ahern et al. 2015; Aybar and Ficici 2009; Conn et al. 2005). Moreover, Alexandridis et al. (2022) and Bereskin et al. (2018) find that larger differences in CSR orientations between acquirers and targets lead to lower announcement returns and long-term operating growth.

Overall, the empirical evidence on the relationship between cultural distance and M&A performance is mixed. However, recent studies (e.g., Ahern et al. 2015; Alexandridis et al. 2022; Bereskin et al. 2018) and meta-analyses mainly support the cultural friction hypothesis

(Rottig 2017; Stahl and Voigt 2008). Furthermore, the capital market shows skepticism towards culturally distant M&As (Chatterjee et al. 1992). Consequently, we argue that the potential benefits of learning and resource recombination are outweighed by friction between culturally distant acquirers and targets during the (post-) merger phase. Therefore, we posit a negative effect of cultural differences on both short-term stock market reactions and post-integration synergy gains. This leads to the following hypotheses:

H1a: **Cultural differences between acquirer and target firms are negatively associated with capital market reactions to M&A announcements.**

H1b: **Cultural differences between acquirer and target firms are negatively associated with post-integration M&A performance.**

2.2 Cultural distance and acquisition premiums

Cultural differences not only affect post-merger outcomes, but they also have a significant impact on the pre-integration phase (Renneboog and Vansteenkiste 2019). In particular, this is reflected in the acquisition premium the acquirer is willing to pay for the target. In M&A price negotiations, target firms often seek the highest possible purchase price. At the same time, acquirers are usually only willing to pay the maximum sum of the target's standalone value and expected synergies (McNichols and Stubben 2015). In addition, the purchase price is directly related to the synergies that the acquiring firm needs to realize to declare the transaction successful (Schweiger 2002; Sirower 1997). Thus, the acquisition premium is considered an important factor in the success of M&As (Haunschild 1994).

The meta-analysis by King et al. (2021) shows a negative correlation between a high acquisition premium and the acquirer's short-term stock performance. This finding is consistent with the hubris hypothesis (Roll 1986), which suggests that acquiring firms tend to be overly optimistic about expected synergies. Moreover, prior research suggests that acquisition premiums are often driven by subjective rather than objective factors (Aktas et al. 2011; Chatterjee and Hambrick 2011; Jentner and Lewellen 2015). Regarding national cultural distance, Lim et al. (2016) highlight that the level of acquisition premiums depends on national differences between the home countries of the acquirer and the target. Similarly, Li and Haleblian (2022) show that the manifestation of different cultural dimensions at the national level (e.g., uncertainty avoidance) affects acquisition premiums.

We argue that the organizational cultural distance between the acquirer and the target leads to information asymmetries between the acquirer and the target. These asymmetries affect the acquisition premium and increase the risk of overpayment by the acquirer (Datta and Puia 1995). Furthermore, we argue that cultural distance prevents the acquirer from accurately assessing the prevailing cultural values of the target firm, leading to a biased assessment of potential synergies (Qiao and Wu 2019). In support of this argument, Laamanen (2007) finds that acquiring firms have difficulty valuing targets with high R&D intensity. These information asymmetries increase acquisition premiums. Moreover, cross-border transactions tend to have higher acquisition premiums because acquirers have less reliable information about target firms (Rossi and Volpin 2004).

In summary, we expect that cultural distance leads to increased information asymmetries, resulting in potential complications in correctly assessing the target's prevailing values and norms. Consequently, we expect that acquirers tend to overestimate the synergy potential of culturally distant targets, leading to potential overpayments. We propose the following hypothesis:

> ***H2*: Cultural differences between acquirer and target firms are positively associated with acquisition premiums.**

2.3 Cultural distance and deal duration

Cultural differences can cause friction during the negotiation phase of an M&A transaction (Dikova et al. 2010). They can affect trust between employees of the acquiring and target firms (Sitkin and Stickel 1996) and impede the flow of information between organizations (Akerlof 1997; Arrow 1974). These factors can lead to higher levels of uncertainty, resulting in longer deal completion times (Nguyen and Phan 2017). However, the empirical research literature provides conflicting evidence on this relationship.

For example, Lawrence et al. (2021) find no significant impact of country-level cultural and institutional factors on deal duration. They argue that acquirers avoid initiating deals in target countries with high cultural and institutional differences. The authors suggest that managerial due diligence has already occurred during the deal initiation phase. Therefore, cultural differences tend to affect the initial selection of a potential target rather than deal duration. On the other hand, Alexandridis et al. (2022) support the notion that cultural distance can increase friction by finding that differences in CSR between merging firms prolong the time required to complete the deal. Similarly, Dikova et al. (2010) find that transactions involving greater differences in cultural and regulatory institutions between the acquiring and target countries experience hold-ups, leading to longer deal durations.

These mixed results highlight the complex relationship between cultural differences and deal duration. Lawrence et al. (2021) suggest that cultural differences may not significantly affect deal duration due to early screening processes. On the contrary, we argue that cultural distance increases the likelihood of protracted negotiations and hold-ups, resulting in longer deal duration. Hence, we propose the following hypothesis:

***H3*: Cultural differences between acquirer and target firms are positively associated with deal durations.**

3 Sample Selection

We obtain deal-specific data from the Securities Data Company (SDC), firm-specific data from Thomson Reuters Refinitiv, and data for the cultural distance proxy is inferred from the text sections of firms' Glassdoor.com reviews. Glassdoor enables employees to anonymously review their employers by providing 5-star ratings for the categories *Overall Rating, Career Opportunities, Compensation & Benefits, Work/Life Balance, Diversity & Inclusion, Senior Management, and Culture & Values*. Furthermore, reviewers are obligated to provide pros and contras about their employer as text inputs ("Share some of the best reasons [downsides] to work at …"). To ensure high review quality, Glassdoor employs a "give to get" policy since new users must provide a review or salary information to access reviews of others. Furthermore, Glassdoor's 5-star ratings are approximately normally distributed, indicating that the reviews are not prone to a response bias (Chemmanur et al. 2019).

To obtain our final deal sample, we first gather an initial sample of 32,330 deals from SDC. Similar to recent M&A literature (e.g., Ahmed et al. 2023; Bena and Li 2014; Bereskin et al. 2018), we require that 1) the status of the deal is completed, 2) the deal value exceeds $1 million, 3) both firms are publicly traded before the deal and the acquirer is publicly traded after the deal, 4) the acquirer owned less than 50% of the target's shares before the deal and more than 90% after the deal, 5) the acquirer does not operate in the investment banking & investment services industry. We also restrict our sample to deals between companies headquartered in the United States, Canada, Australia, or the United Kingdom to ensure a sufficiently large number of English reviews on Glassdoor. These requirements reduce the initial sample to 4,558 deals.

To determine whether a potential deal has its acquirer and target listed on Glassdoor, we automatically extract potentially matching Glassdoor links from three search engine queries. We then match Glassdoor and SDC deal data by applying fuzzy string matching between firm names in the SDC data and the extracted Glassdoor links. For cosine similarity values below 0.8, we manually verify that the correct Glassdoor link was extracted by comparing information on the firm's industry, year founded, location, name changes, and the firm's website. For our final sample, we require at least 10 Glassdoor reviews for both the acquirer and target in the years prior to the deal announcement (similar to Campbell and Shang 2021). Finally, after removing observations with missing values, our final sample contains 243 deals with 345,305 acquirer and 67,814 target reviews from 439 unique firms.

4 Identification Strategy

To test our hypotheses, we apply the following cross-sectional two-way fixed effects regression models to our event-study setting:

$$DealOutcome_m = \alpha + \beta_1 CulturalDistance_{i,j} + \sum \beta_m DealControls_m + \sum \beta_i AcquirorControls_i + \sum \beta_j TargetControls_k + \gamma_i + \delta_m + \varepsilon_m \quad (1)$$

where DealOutcome_m represents the respective dependent variable of deal *m*, *Cultural-Distance* between firm *i* and *j* represents our main independent variable, and *Controls* are three vectors containing several control variables of deal *m*, acquirer firm *i*, and target firm *j*. All variables are described in detail below, and Appendix B provides a summary of the variables. To reduce potential endogeneity concerns, we introduce acquirer industry fixed effects (γ_i) to control for time-invariant industry characteristics (similar to Bereskin et al. 2018; Suk and Wang 2021). We also apply time fixed effects (δ_t) to control for merger waves and macroeconomic trends.[1] Standard errors are robust to heteroskedasticity and serial correlations within 2-digit SIC industry clusters of the acquirers.[2]

[1] In models analyzing effects around the announcement date (e.g., announcement cumulative abnormal returns and acquisition premiums), we use the year when the deal was publicly announced for the first time by an involved party. In all other deal outcome models, the year fixed effects refer to the year in which the entire deal was completed.

[2] As a robustness check, we also cluster standard errors at the acquirer level.

4.1 Cultural distance measure

Previous research mostly examines the organizational cultural distance between the acquirer and the target using (small-scale) survey studies (Rottig 2017). More recently, scholars have started to use machine learning techniques to measure corporate culture (e.g., Corritore et al. 2020; Li et al. 2021). In this vein, Latent Dirichlet Allocation (LDA) is often used for unsupervised topic modeling (Blei et al. 2001). LDA assigns topic probability scores to each document based on word co-occurrences. Due to its unsupervised approach, a caveat of using LDA is that the number of topics to be discovered has to be set a-priori (Bochkay et al. 2023). Thus, topic categories must be determined manually after running the model, leaving leeway for topic overlap and potential misinterpretation. Another (more advanced) approach to measuring corporate culture is based on the word2vec algorithm (Mikolov et al. 2013). Li et al. (2021) use word2vec as a semi-supervised word embedding model to generate a context-specific dictionary to measure corporate culture.

Nevertheless, LDA and word2vec underperform in natural language processing tasks compared to state-of-the-art transformer models (Bochkay et al. 2023; Koch and Pasch 2022; Mosel et al. 2023). These large language models outperform LDA and word2vec for several reasons. First, transformer models use the principle of attention (Vaswani et al. 2017), which enables transformer models to learn contextual representations. Thus, they better understand short, nuanced, or colloquial expressions because they recognize the context of whole sentences or statements. In contrast, LDA's bag-of-words approach ignores the immediate context of words. word2vec is also not able to fully reflect the context of a document. Second, transformers can handle ironic expressions. Third, they can handle statements that have opposite meanings, even in the absence of a negation.

Due to these advantages, we use the pre-trained CultureBERT large language model (Koch and Pasch 2022) as the basis for our cultural distance scores. For example, CultureBERT correctly classifies the phrase "they are quick to throw you under the bus" (p. 12) as a statement contradicting the clan culture. Furthermore, CultureBERT does not assign a competitive culture to the review phrase "the job offers a competitive compensation" (p. 13). CultureBERT builds on the widely used RoBERTa large language model (Liu et al. 2019). To adequately classify cultural dimensions, CultureBERT was manually fine-tuned for the four culture dimensions of the Competing Values Framework (CVF; Cameron et al. 2006; Quinn and Rohrbaugh 1983). Fine-tuning was accomplished by manually labeling 2,000 Glassdoor reviews. During this process, multiple people classified which of the four culture dimensions (clan, adhocracy, market, hierarchy) best fit the overall tone of a review (see Appendix A for Glassdoor review examples associated with the four dimensions).

To determine the organizational culture score of each review, we merge the pro and contra text sections of each Glassdoor review into a corpus of text and then tokenize them. Since the performance of RoBERTa declines with increasing token length (Koch and Pasch 2022), we limit the number of words per review to 300, with the median length of our reviews being 128 words. Next, we use CultureBERT to estimate a probability score for each tokenized review, resulting in four scores between 0 and 1. Larger values indicate a higher affiliation to the respective dimension. We then calculate an annual average for each organizational culture dimension across all company reviews in the given year. Next, we average the annual culture scores up to the announcement year of each deal.

Finally, we determine the organizational cultural distance between the acquirer and target by calculating the Euclidean distance of the acquirer *i* and target *j* between the four culture categories *c* (similar to Kogut and Singh 1988):

$$CulturalDistance_{i,j} = \sqrt{\sum_{c=1}^{4}(acq\_dimension\_culture_c - tar\_dimension\_culture_c)^2} \quad (2)$$

In addition to using *deal_culturaldistance* as the main independent variable, we also run models with the four absolute differences of the organizational cultural dimensions as separate independent variables. This allows us to observe the direct impact of each dimension on the respective merger outcome. Using separate culture measures is especially helpful in cases where the dimensions have opposing effects on the respective outcome.

4.2 Dependent measures

Capital market reactions and post-merger synergies

Short-term event studies are by far the most popular approach to evaluate M&A success (Renneboog and Vansteenkiste 2019). Thus, we use acquirer and combined (value-weighted acquirer and target) announcement cumulative abnormal returns (CARs) to measure capital market reactions to deal announcements. We follow the established event study methodology to estimate the market model over the period of 50 to 250 days prior to the announcement date (e.g., Brown and Warner 1985; Suk and Wang 2021). We aggregate acquirers' daily abnormal returns over [-2, 2] and [-5, 5] days around the announcement date (*acq_car2*, *acq_car5*). To calculate the combined CAR of the M&A pair (*deal_totalcar_weighted*), we multiply the CAR of the acquirer and the target by their relative market capitalization.

Since CARs potentially understate the bidder's true acquisition gains, we employ an accounting-based proxy to measure long-term synergy gains (Barraclough et al. 2013; Renneboog and Vansteenkiste 2019). Consistent with previous studies on post-merger performance (e.g., Morosini et al. 1998; Suk and Wang 2021), we use return on sales growth as a proxy (*acq_ros_1year, 3years*) for long-term synergistic gains. We measure long-term performance as the change in net sales divided by assets over the two years after deal completion. Compared to return on assets, return on sales growth is calculated from cash flows, making it robust to inflation and possible differences in accounting standards. As a result, it is likely to reflect the long-term success of the transaction better than the return on assets (Zhu et al. 2019).

Transaction phase outcomes

As proxies for transaction phase outcomes, we use acquisition premiums and total deal duration (e.g., Bereskin et al. 2018; Lawrence et al. 2021; Lee et al. 2019; Recendes et al. 2022). We measure acquisition premiums as the difference between the acquirer's payment and the target's market value, divided by the target's market value one month before the deal announcement (*deal_premium_1month*) (Lee et al. 2019). Prior research uses extended periods to mitigate the potential impact of information leakage immediately before the announcement (Reuer et al. 2012). Hence, we measure the target's market value at three increasing points in time (one day; one week; one month) prior to the deal announcement. *deal_duration_ln* is the natural logarithm of the difference in days between the date the deal was completed and the date when the deal was announced for the first time by an involved party.

4.3 Control variables

Similar to Ahmed et al. (2023), Bereskin et al. (2018), and Suk and Wang (2021), we add a battery of control variables to our regression models that cover deal, acquirer, and target characteristics that may affect (post-) merger phase outcomes (for a detailed variable list and description, see Appendix B).

Deal controls

Bereskin et al. (2018) find that larger deals are associated with lower acquirer CARs. In addition, larger deal values are associated with higher deal complexity, which increases deal duration (Lawrence et al. 2021). Therefore, we include the natural logarithm (to mitigate skewness) of the deal value (*deal_value_ln*) as a control. Next, we include two dummy variables indicating whether the deal was all cash (*deal_all_cash_dummy*) or all equity financed (*deal_all_stock_dummy*), as Loughran and Vijh (1997) find that equity-financed acquisitions generate lower returns than cash-financed ones. In addition, deal financing affects acquisition premiums (Ghosh and Ruland 1998).

We also control for whether the deal involves a tender offer (*deal_tenderoffer_dummy*), as Chen et al. (2018) find that tender offers are positively associated with acquisition synergies. However, tender offers could also lead to higher acquisition premiums, as the target's management may initially resist the tender offer (Raghavendra and Vermaelen 1998). Moreover, Schwert (2000) finds that acquirers earn lower abnormal returns in hostile deals. Thus, we include *deal_friendly_dummy*. *deal_relatedness_dummy* captures whether the acquirer and the target operate in the same two-digit SIC code since industry familiarity potentially reduces deal uncertainties (Morck et al. 1990).

Since our sample includes international deals, we control for national cultural distances between acquirers' and targets' home countries using Kogut and Singh's (1988) measure of cultural distance. It is calculated as the Euclidean distance between Hofstede's six cultural dimensions individualism, power distance, uncertainty avoidance, femininity, indulgence, and long-term orientation (Hofstede 2001; Lawrence et al. 2021). For merger phase analyses (acquisition premiums and deal duration), we also include the number of bidders (*deal_bidders*), as the presence of multiple bidders may increase the bargaining power of the target. This potentially increases deal durations and acquisition premiums (Giliberto and Varaiya 1989).[3]

Acquirer and target controls

Moeller et al. (2004) observe that the acquirer's size has a negative impact on its announcement returns due to an increased likelihood of engaging in value-destroying mergers induced by management entrenchment. Consequently, we control for the natural logarithm of the acquirer's size (*acq_assets_lastyear_ln)*. In this sense, we also include the target's size (*tar_assets_lastyear_ln)*, which directly affects the target's attractiveness and deal outcomes (Chen et al. 2018; Lee et al. 2019). We also include the acquirer's operating performance

[3] We replace four missing *deal_bidders* values with one (the median value).

(*acq_roa_lastyear*), as Morck et al. (1990) find that firms with higher operating performances are more successful acquirers. Similarly, we also control for the target's operating performance (*tar_roa_lastyear*) due to its potential positive impact on post-merger synergies. Higher target profitability also increases target attractiveness, which potentially increases acquisition premiums (Hayward and Hambrick 1997).

To control for the acquirer's general industry performance, we add the acquirer's industry growth during the 12 months before the deal announcement (Ellis et al. 2011) in addition to the industry fixed effects (*acq_industrygrowth*). Additionally, we control for the target's market-to-book ratio (*tar_mbratio*) because a high market valuation makes it more difficult to realize growth opportunities after deal completion (Laamanen 2007). Finally, Zollo and Singh (2004) find that the acquirer's prior deal experience positively impacts acquisition performance. Therefore, we include the acquirer's prior deal experience in the last three years as another control (*acq_dealexperience*). Since firms with a longer history might have higher (deal) experience, we also include the age of the acquirer and the target as controls (Naranjo-Valencia et al. 2011).

5 Empirical Results

5.1 Descriptive statistics

Table 1 shows the number of deals by announcement year. Since Glassdoor was launched in 2008 and firms only slowly began to receive reviews, most deals in our sample occurred after 2011.

[Insert Table 1 about here]

Table 2 shows the descriptive statistics for all the variables used in our specified models. The four cultural affiliations of the firms (e.g., *acq_clan_culture*), represented by probability scores between 0 and 1, have their minimum (maximum) values for each dimension below the threshold of 0.08 (above the threshold of 0.51). The means of each dimension are comparable for acquirers and targets. However, the standard deviations are lower for the acquirer dimensions because there are approximately five times more acquirer Glassdoor reviews available than target reviews. The average absolute difference between the acquirer's and the target's cultural dimensions is around 0.09. In 57% of the deals, the culture with the highest probability value differs between the acquirer and target (*deal_dominant_*culture_*diff*). In 35% (41%) of the deals, the acquirers' (targets') most dominant culture is reflected by the market dimension (*acq_dominant_market*). Our main independent variable *deal_culturaldistance,* representing the Euclidean distance between the four cultural dimensions of the acquirer and the target, ranges from 0.029 to 0.58, with a mean of 0.201.

[Insert Table 2 about here]

5.2 Results of H1 – The impact of cultural distance on capital market reactions and post-merger synergies

For H1a, we argue that for culturally distant M&A pairs, the cultural friction hypothesis leads to pessimistic capital market reactions in the form of lower announcement returns. Models 1-3 of Table 3 use the Euclidean distance between the acquirer's and target's four CVF culture dimensions (clan, adhocracy, market, hierarchy) as our main independent variable (*deal_culturaldistance*). In Model 1 (2), we center the acquirer's CAR [-2,2] ([-5,5]) days around the deal announcement. We also use the [-5,5] combined (value-weighted) CAR of the acquirer and target as an additional measure of capital market reactions (Model 3). The results show a negative, however, insignificant relationship between the cultural distance between the acquirer and the target and the acquirer's respective CAR ($t < -1.15$, $p < 0.26$). The coefficients of the control variables are generally consistent with the expected directions suggested by the prior literature. For example, *deal_value_ln* has a positive coefficient (Suk and Wang 2021), *deal_all_cash_dummy* has a positive coefficient (Alexandridis et al. 2022; Bereskin et al. 2018; Chakrabarti et al. 2009), and *tar_mbratio* has a negative coefficient (Laamanen 2007).

[Insert Table 3 about here]

The results of the aggregated cultural distance measure do not corroborate the cultural friction hypothesis. We therefore separately analyze the impact of the four CVF culture dimensions on capital market reactions. Thus, we replace the *deal_culturaldistance* variable with four separate variables (Models 4-6). Each variable captures the absolute difference between the acquirer's and target's attitude towards the respective CVF culture dimension. Results of Models 4-6 indicate that the capital market expects synergy losses of acquirers and targets that have large differences in their market orientations ($t < -2.13$, $p < 0.039$), corroborating H1a.

The market dimension is reflected by publicly available information such as market share and profitability (Cameron et al. 2006). Thus, we argue that differences in firms' attitudes toward a market orientation are more easily observable by outsiders. In contrast, the other three dimensions are more nuanced and, thus, more difficult for other market participants to observe. In addition, the market dimension is the most dominant in more than a third of the deals, with the (second) highest standard deviation of the four dimensions (see Table 2). Lastly, the results of the differences in the market orientation (Models 4-6) corroborate the cultural friction hypothesis since higher market differences decrease announcement returns.

For H1b, we analyze whether the synergistic benefits from culturally similar acquirers and targets materialize by measuring post-merger synergies (Table 4). First, we replace the CAR dependent variable with the acquirer's growth in return on sales in the year (three years) following the deal. Even though the effect of the *deal_culturaldistance* coefficients is increasing from one to three years, the coefficients of Models 1 ($t = 0.15$, $p = 0.88$) and 2 ($t = 0.71$, p

= 0.48) are positive but insignificant. Thus, we cannot accept the alternative hypothesis of H1b.[4]

Next, we analyze the four separate culture dimensions (Models 3 and 4). In contrast to the capital market reactions to the differences in the market orientations of M&A pairs, Model 4 shows that differences in clan orientations lead to significant decreases in post-merger return on sales growth (t = -1.82, p = 0.076). The result shows that differences in the emphasis on collaboration, teamwork, and employee development (clan orientation) of M&A pairs decrease the acquirer's post-merger operating growth, partially corroborating H1b.

[Insert Table 4 about here]

5.3 Results of H2 – The impact of cultural distance on acquisition premiums

In H2, we argue that cultural distance between the acquirer and the target reduces the acquirer's ability to accurately assess the target's prevailing values and norms concerning synergy potential. As a result, the acquirer is more likely to misperceive the target's true value. Table 5 shows the regression results with acquisition premiums as the dependent variable. Model 1 shows a significant positive relationship between cultural distance and acquisition premiums (t = 1.75, p = 0.087). Thus, firms that acquire culturally distant targets overestimate the potential synergy realizations and thus pay higher acquisition premiums.

[Insert Table 5 about here]

To derive which cultural dimensions specifically lead to these overpayments, we rerun the analyses using the four separate CVF cultural difference variables (Model 2). Acquisition premiums significantly increase for acquirers and targets exhibiting different market orientations (t = 2.56, p = 0.014). Similar to previous research (Lee et al. 2019; Reuer et al. 2012),

[4] Results of Models 3 and 4 show that two dimensions have a positive coefficient (market and hierarchy), while the other two (clan and adhocracy) have a negative coefficient. As a result, the effect of the aggregated cultural distance variable in Models 1 and 2 is most likely rendered insignificant.

we use multiple pre-announcement periods to measure the target's market value. In doing so, we increase robustness by mitigating the potential impact of information leakage immediately before the announcement. Irrespective of using a 1-day, 1-week, or 1-month pre-announcement period, acquisition premiums increase significantly for M&A pairs with different market orientations ($p < 0.047$, untabulated).

To further substantiate this effect, we conduct an additional analysis to examine whether overpayment is driven by deals in which the acquirer has a greater market orientation than the target. Therefore, we replace the independent variables for the cultural dimensions in Model 2 with two binary variables (similar to Bereskin et al. 2018). First, we calculate the (non-absolute) difference between the acquirer's and the target's market orientation. Next, we rank the differences between all deals into quartiles. The binary variable *deal_market_diff_acq_high* (*low*) takes a value of one if the market difference between the acquirer and target is ranked in the highest (lowest) quartile. The results of Model 3 show that overpayments are caused by deals in which the acquirer has a significantly lower market orientation than its target ($t = 2.48$, $p = 0.017$). A possible explanation for this effect is that acquirers with a low market orientation overestimate the positive synergies of acquiring a target with a high market orientation.

5.4 Results of H3 – The impact of cultural distance on deal duration

Prior research finds mixed evidence concerning the impact of cultural distance on deal duration. Lawrence et al. (2021) find that cultural and institutional factors do not affect deal duration, suggesting that managerial due diligence has already occurred during the deal initiation stage. In contrast, Alexandridis et al. (2022) highlight the potential for protracted negotiations and hold-ups. Table 6 Model 1 supports the notion of Lawrence et al. (2021), as the overall effect of cultural distance on *deal_duration_ln* is insignificant (t = -0.80, p = 0.428). However, Model 2 shows that larger absolute differences in the hierarchy dimension decrease deal duration (t = -1.11, p = 0.055).[5]

[Insert Table 6 about here]

In Model 3, we conduct an additional analysis to examine whether specific hierarchical differences lead to shorter deal durations. To do so, we again replace the independent variable with two binary variables, as in the analyses of Table 6. The binary variable *deal_hierarchy_diff_acq_high* (*low*) takes a value of one if the hierarchical difference between the acquirer and the target is in the highest (lowest) quartile. The coefficient of *deal_hierarchy_diff_acq_high* is significantly negative (t = -2.72, p < 0.01). This result shows that when the acquirer has a significantly higher hierarchy score than its target, the deal is completed significantly faster. Since the hierarchical culture emphasizes clear rules and guidelines, we argue that there is less leeway for time-consuming debates induced by the target, allowing for a faster deal completion time. Thereby, the acquirer and target can spend more time on post-merger integration processes.

5.5 Robustness checks

We run several additional robustness checks to validate our results. First, we replace our main independent variable *deal_culturaldistance,* which is calculated using the Euclidean distance between the four Competing Values Framework dimensions. Corritore et al. (2020) use the Jensen-Shannon (JS) divergence to derive their cultural homogeneity measure between Glassdoor reviews. The JS measures the difference between probability distributions by con-

[5] Since differences in *deal_market_culture_abs* also have an (almost significant) effect in the opposite direction (t = 1.45, p = 0.153) compared to effect of absolute differences in the hierarchy dimension, we argue that total effect of *deal_culturaldistance* on deal duration is rendered insignificant by these opposing effects.

sidering the relative probabilities of different categories. As a result, the JS produces a divergence value that indicates the difference in distributions. Even though the *deal_culturaldistance_js* has a lower mean and standard deviation than the *deal_culturaldistance* (see Table 2), the *deal_culturaldistance_js* coefficients of all main analyses remain unchanged (untabulated), except that the coefficient of *deal_culturaldistance_js* of Table 1 Model 2 (*acq_car5*) becomes significant (p = 0.084), which corroborates the cultural friction hypothesis (H1a).

Second, we cluster standard errors on the acquirer's 2-digit SIC industry level (similar to Bereskin et al. 2018). Since 33 acquirers conducted multiple deals in our sample, we also re-run our main analyses with standard errors clustered at the acquirer level (similar to Ahmed et al. 2023). Results show that significance levels of the culture coefficients do not change when clustering standard errors at the acquirer level (untabulated).

Third, our sample contains five observations in which the acquirer already held more than 0% and less than 50% of the target's shares. To ensure that the acquirer did not already influence the target's culture, we remove these toehold deals from our main analyses. The significance levels of the culture coefficients (except in Table 5 Model 1, p = 0.187) remain unchanged for all primary analyses (untabulated).

6 Conclusion

Using a novel sample of 243 M&As between 2008 and 2021, this study provides evidence on the impact of differences in the organizational cultures of acquirers and targets on M&A outcomes. We apply state-of-the-art deep learning (using the large language model CultureBERT; Koch and Pasch 2022) to about 400,000 Glassdoor employee reviews to infer our cultural distance score. Thereby, we directly address weaknesses of previous studies that often use subjective measures to analyze cultural differences and M&A success, which reduce the comparability and objectivity of the results (Rottig 2017).

Our analyses provide several insights. First, we show that differences in market and clan culture dimensions of M&A pairs negatively impact capital market reactions and post-merger synergies, partially supporting H1a and H1b. Second, our results show that cultural differences increase acquisition premiums since cultural differences negatively affect the acquirer's ability to determine the target's true value. Primarily, this phenomenon arises from acquirers with a limited market orientation who tend to overestimate the synergistic benefits of acquiring a target with a strong market orientation. Third, we provide insights into the scarcely studied relationship between organizational cultural distance and acquisition premiums. Our results reveal that cultural distance does not prolong a deal's completion speed. Nevertheless, we find that deal completion is significantly faster when the acquirer possesses a significantly higher hierarchical orientation than the target due to a lower likelihood of the target causing time-consuming debates. Thereby, acquirers and targets can spend more time focusing on the post-merger integration phase.

Future research could examine whether cultural distance between M&A pairs fosters or mitigates innovation growth (Büschgens et al. 2013). Even though contradicting the cultural friction hypothesis, cultural distance could foster innovation growth since Corritore et al. (2020) find that intrapersonal cultural heterogeneity within firms leads to larger patent outputs. Prior research only examined the general relationship between M&A behavior and corporate innovation (Bena and Li 2014). Thus, future research could provide further insights by considering the acquirer's and target's corporate culture when examining the impact of M&As on corporate innovation.

Our study adds to the M&A literature in several ways by providing evidence of the relationship between cultural distance and M&A success, acquisition premiums, and deal duration. Nevertheless, the study also comes with limitations. Despite its widespread use in the M&A literature (Renneboog and Vansteenkiste 2019), it should be noted that the study of short-term M&A success assumes that the market reaction to M&A announcements is (almost) instantaneous, complete, and unbiased, which requires at least the semi-strong form of the efficient market hypothesis (Fama 1970). Although evidence indicates that markets consider cultural differences when evaluating transactions (Aktas et al., 2011), it is plausible that this information is misinterpreted or that other factors confound the impact of cultural distance.

Moreover, we only examine M&A transactions between firms in major economies with English as their primary language (North America, Canada, Australia, and the United Kingdom). This limits the transferability of our findings to other regions. Future research could achieve broader geographic coverage by adjusting the research design to obtain more representative results, e.g., by using employee reviews of local review websites. Lastly, we only examine completed M&A transactions. Future research could examine whether culturally distant M&A pairs are potentially more likely to withdraw from announced deals.

Overall, our findings have economic relevance for practitioners. Thus, we suggest that acquiring firms should carefully assess differences between their and the target's cultural orientation during and before engaging in M&A transactions. Otherwise, potential negative (monetary) consequences can occur, and potential synergies might be lost.

Tables

Table 1 – Deals by announcement year

Announcement year	Deals	Percentage	Total percentage
2008	2	0.82	0.82
2009	5	2.06	2.88
2010	3	1.23	4.11
2011	4	1.65	5.76
2012	11	4.53	10.29
2013	10	4.12	14.41
2014	15	6.17	20.58
2015	36	14.81	35.39
2016	36	14.81	50.2
2017	22	9.05	59.25
2018	22	9.05	68.3
2019	23	9.47	77.77
2020	21	8.64	86.41
2021	33	13.58	100.0
Total	**243**	**100**	**100**

This table presents the number of deals per year. The descriptive statistics are based on 243 deal observations for the years 2008 to 2021.

Table 2 – Descriptive statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Firm Culture Affiliation					
acq_clan_culture	243	.259	.084	.036	.511
acq_adhocracy_culture	243	.203	.104	.013	.626
acq_market_culture	243	.276	.098	.077	.58
acq_hierarchy_culture	243	.262	.083	.051	.516
tar_clan_culture	243	.257	.1	.025	.57
tar_adhocracy_culture	243	.22	.117	.036	.691
tar_market_culture	243	.293	.117	.038	.603
tar_hierarchy_culture	243	.259	.084	.036	.511
acq_dominant_adhocracy	243	.14	.348	0	1
acq_dominant_clan	243	.214	.411	0	1
acq_dominant_hierarchy	243	.3	.459	0	1
acq_dominant_market	243	.346	.477	0	1
tar_dominant_adhocracy	243	.198	.399	0	1
tar_dominant_clan	243	.21	.408	0	1
tar_dominant_hierarchy	243	.185	.389	0	1
tar_dominant_market	243	.407	.492	0	1
Deal Cultural Distance					
deal_clan_culture_abs	243	.083	.068	.001	.404
deal_adhocracy_culture_abs	243	.079	.075	0	.433
deal_market_culture_abs	243	.093	.075	0	.419
deal_hierarchy_culture_abs	243	.091	.069	.001	.333
deal_dominant_culture_diff	243	.572	.496	0	1
deal_culturaldistance	243	.201	.103	.029	.58
deal_culturaldistance_js	243	.144	.072	.023	.401
Dependent Variables					
acq_car2	243	-.017	.158	-1.826	.33
acq_car5	243	-.035	.313	-3.598	.705
deal_car_weighted	243	-.016	.135	-1.634	.318
acq_ros_1year	241	-.039	1.042	-15.531	3.135

acq_ros_2years	187	.015	.131	-.571	1.394
deal_premium_1day	226	.328	.306	-.312	2.367
deal_premium_1week	226	.356	.313	-.307	2.438
deal_premium_1month	226	.38	.306	-.341	2.625
deal_duration_ln	241	4.781	.717	3.332	6.413
Deal Controls					
deal_value_ln	243	21.568	1.672	17.272	25.156
deal_hofstede_distance	243	.031	.081	0	.345
deal_bidders	243	1.070	.286	1	3
deal_relatedness_dummy	243	21.568	1.672	17.272	25.156
deal_all_cash_dummy	243	.152	.36	0	1
deal_all_stock_dummy	243	.181	.386	0	1
deal_tenderoffer_dummy	243	.992	.091	0	1
deal_friendly_dummy	243	.687	.465	0	1
Acquirer Controls					
acq_age	243	69.984	49.831	6	232
acq_assets_lastyear_ln	243	23.081	1.773	18.594	27.496
acq_roa_lastyear	243	.419	.946	-.33	10.077
acq_industrygrowth	243	.009	.051	-.037	.551
acq_dealexperience	243	1.535	.937	1	6
Target Controls					
tar_age	243	52.119	40.36	7	191
tar_assets_lastyear_ln	243	21.209	1.847	16.507	27.386
tar_roa_lastyear	243	.017	.163	-.511	2.028
tar_mbratio	243	2.363	4.503	0	59.87

The descriptive statistics are based on 243 deal observations for the years 2008 to 2021. A detailed variable description can be found in Appendix B.

Table 3
H1a: Capital market reactions (CARs)

	(1) acq_car [-2,2]	(2) acq_car [-5,5]	(3) deal_car_weighted [-5,5]	(4) acq_car [-2,2]	(5) acq_car [-5,5]	(6) deal_car_weighted [-5,5]
deal_culturaldistance	-.128	-.322	-.096			
	(-1.153)	(-1.495)	(-1.189)			
deal_clan_culture_abs				-.136	-.288	-.058
				(-1.128)	(-1.175)	(-.621)
deal_adhocracy_culture_abs				.005	.012	.015
				(.051)	(.06)	(.185)
deal_market_culture_abs				-.262**	-.485**	-.236**
				(-2.175)	(-2.126)	(-2.325)
deal_hierarchy_culture_abs				.167	.119	.139
				(1.201)	(.545)	(1.169)
deal_value_ln	.037*	.074**	.032*	.041*	.079**	.034*
	(1.834)	(2.088)	(1.929)	(1.939)	(2.157)	(1.99)
deal_hofstede_distance	-.029	.084	-.028	-.027	.079	-.026
	(-.271)	(.45)	(-.307)	(-.298)	(.478)	(-.334)
deal_all_cash_dummy	.008	.031	.006	.004	.026	.003
	(.273)	(.671)	(.277)	(.146)	(.578)	(.146)
deal_all_stock_dummy	-.016	-.043	-.021	-.014	-.043	-.02
	(-.457)	(-.646)	(-.667)	(-.414)	(-.644)	(-.643)
deal_tenderoffer_dummy	-.012	-.021	-.015	-.01	-.019	-.014
	(-.483)	(-.472)	(-.725)	(-.417)	(-.422)	(-.664)
deal_friendly_dummy	-.008	-.05	-.002	.044	.013	.04
	(-.151)	(-.448)	(-.043)	(.655)	(.104)	(.718)
deal_relatedness_dummy	-.001	.001	-.004	.003	.006	-.001
	(-.038)	(.012)	(-.199)	(.132)	(.132)	(-.025)
acq_age	.001	.001	.001	.001	.001	.001

	(1.567)	(1.458)	(1.487)	(1.617)	(1.481)	(1.506)
acq_assets_lastyear_ln	.001	.002	-.002	-.001	.001	-.002
	(.049)	(.144)	(-.225)	(-.019)	(.059)	(-.295)
acq_roa_lastyear	-.027	-.055	-.025	-.026	-.054	-.025
	(-1.432)	(-1.635)	(-1.649)	(-1.403)	(-1.639)	(-1.627)
acq_industrygrowth	.799**	1.403**	.697**	.866**	1.51**	.73**
	(2.272)	(2.211)	(2.2)	(2.447)	(2.332)	(2.3)
acq_dealexperience	.003	-.006	.005	-.001	-.011	.002
	(.203)	(-.197)	(.394)	(-.031)	(-.37)	(.199)
tar_age	-.001	-.001	-.001	-.001	-.001	-.001
	(-1.608)	(-1.675)	(-1.614)	(-1.58)	(-1.636)	(-1.586)
tar_assets_lastyear_ln	-.037	-.076	-.032	-.038	-.077	-.031
	(-1.369)	(-1.56)	(-1.415)	(-1.361)	(-1.546)	(-1.374)
tar_roa_lastyear	-.067	-.158	-.075	-.055	-.136	-.066
	(-.786)	(-.884)	(-.933)	(-.63)	(-.737)	(-.804)
tar_mbratio	-.002	-.004	-.002	-.003	-.005	-.003
	(-.966)	(-.828)	(-.929)	(-1.144)	(-.971)	(-1.11)
Observations	243	243	243	243	243	243
Adj R^2	.15	.182	.181	.153	.178	.183
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

This table presents the effect of cultural distance between acquirers and targets on deal announcement cumulative abnormal returns (CARs). The dependent variable for Models 1, 2, 4 and 5 is the acquirer's cumulative abnormal return, with the day range being centered around the announcement given in brackets. The dependent variable for Models 3 and 6 is the value-weighted cumulative abnormal return of the acquirer and target. The main independent variable in Models 1-3 is *deal_cultural-distance*, with higher values indicating higher cultural dissimilarities between the acquirer and target. The main independent variables in Model 2 are the absolute cultural differences (in the four Competing Values Framework categories clan, adhocracy, market, hierarchy) between the acquirer and target. A detailed variable description can be found in Appendix B. Constant terms are estimated but not reported. We include the acquirer's industry and year fixed effects in all models. We cluster standard errors at the 2-digit SIC industry level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 4
H1b: Synergy gains

	(1) acq_ros_ 1year	(2) acq_ros_ 3years	(3) acq_ros_ 1year	(4) acq_ros_ 3years
deal_culturaldistance	.06	.087		
	(.149)	(.706)		
deal_clan_culture_abs			-1.24	-.371*
			(-1.439)	(-1.819)
deal_adhocracy_culture_abs			-.331	-.153
			(-.573)	(-.903)
deal_market_culture_abs			2.112	.29
			(1.206)	(1.152)
deal_hierarchy_culture_abs			-1.127	.233
			(-.662)	(1.063)
deal_value_ln	.138	-.022	.142	-.016
	(.858)	(-.693)	(.907)	(-.586)
deal_hofstede_distance	-3.077	-.036	-3.031	-.038
	(-.991)	(-.217)	(-.994)	(-.216)
deal_all_cash_dummy	.05	-.047***	.074	-.052***
	(.371)	(-3.277)	(.463)	(-3.443)

deal_all_stock_dummy	-.14	.059	-.145	.054
	(-.465)	(.536)	(-.487)	(.514)
deal_tenderoffer_dummy	-.015	.001	-.019	.004
	(-.183)	(.023)	(-.245)	(.263)
deal_friendly_dummy	.307	-.078*	.097	-.021
	(.821)	(-1.937)	(.298)	(-.406)
deal_relatedness_dummy	-.299	-.025	-.326	-.023
	(-1.153)	(-1.171)	(-1.187)	(-1.098)
acq_age	.001	-.001	.001	-.001
	(.894)	(-.197)	(.965)	(-.055)
acq_assets_lastyear_ln	-.075	-.008	-.078	-.006
	(-.943)	(-.507)	(-.955)	(-.416)
acq_roa_lastyear	-.046	-.039	-.033	-.032
	(-.721)	(-1.611)	(-.528)	(-1.575)
acq_industrygrowth	-.979	.396	-.758	.616
	(-.426)	(1.299)	(-.308)	(1.545)
acq_dealexperience	.04	-.01	.056	-.015
	(.576)	(-.725)	(.612)	(-1.071)
tar_age	.001	-.001	.001	-.001
	(.609)	(-.37)	(.488)	(-.487)
tar_assets_lastyear_ln	-.078	.032	-.096	.023
	(-.666)	(1.065)	(-.785)	(1.013)
tar_roa_lastyear	-.242	-.004	-.254	-.008
	(-.915)	(-.118)	(-.898)	(-.224)
tar_mbratio	-.003	.001	.002	.001
	(-.37)	(1.461)	(.304)	(1.26)
Observations	241	187	241	187
Within R^2	.076	.095	.087	.118
Industry FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

This table presents the effect of cultural distance between acquirers and targets on synergy gains. The dependent variable for all Models is the acquirer's growth in return on sales in the one and the three years after the deal announcement. The main independent variable in Models 1 and 2 is *deal_culturaldistance*, with higher values indicating higher cultural dissimilarities between the acquirer and target. The main independent variables in Models 3 and 4 are the absolute cultural differences (in the four Competing Values Framework categories clan, adhocracy, market, hierarchy) between the acquirer and target. A detailed variable description can be found in Appendix B. Constant terms are estimated but not reported. We include the acquirer's industry and year fixed effects in all models. We cluster standard errors at the 2-digit SIC industry level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 5
H2: Acquisition premiums

	(1) deal_premium_ 1month	(2) deal_premium_ 1month	(3) deal_premium_ 1month
deal_culturaldistance	.494*		
	(1.746)		
deal_clan_culture_abs		-.161	
		(-.382)	
deal_adhocracy_culture_abs		.304	
		(.888)	
deal_market_culture_abs		.902**	
		(2.56)	
deal_hierarchy_culture_abs		-.108	
		(-.282)	
deal_market_diff_acq_high			.081
			(1.259)
deal_market_diff_acq_low			.146**
			(2.483)
deal_value_ln	-.009	.001	-.012
	(-.383)	(.049)	(-.425)
deal_hofstede_distance	-.134	-.165	-.169
	(-.512)	(-.695)	(-.713)
deal_all_cash_dummy	.107*	.124**	.102*
	(1.763)	(2.051)	(1.793)
deal_all_stock_dummy	-.104	-.097	-.119
	(-1.288)	(-1.097)	(-1.538)
deal_tenderoffer_dummy	-.104	-.104	-.082
	(-1.464)	(-1.392)	(-1.09)
deal_friendly_dummy	.29*	.215	.237*
	(1.961)	(1.314)	(1.872)
deal_relatedness_dummy	.087	.085	.076
	(1.46)	(1.573)	(1.396)
deal_bidders	.136*	.14**	.167***
	(1.99)	(2.034)	(2.783)
acq_age	-.001	-.001	-.001
	(-.777)	(-.637)	(-.939)
acq_assets_lastyear_ln	.037	.031	.036
	(.887)	(.682)	(.907)
acq_roa_lastyear	.001	.002	.001
	(.023)	(.067)	(.011)
acq_industrygrowth	-1.401	-1.427	-1.04
	(-1.509)	(-1.574)	(-1.295)
acq_dealexperience	-.018	-.014	-.012
	(-.458)	(-.408)	(-.346)
tar_age	.001	.001	.001
	(.588)	(.637)	(.669)
tar_assets_lastyear_ln	-.005	-.016	-.007
	(-.149)	(-.497)	(-.226)
tar_roa_lastyear	-.298	-.302	-.447
	(-.968)	(-.917)	(-1.258)
tar_mbratio	-.004**	-.003*	-.004
	(-2.51)	(-1.833)	(-1.61)
Observations	226	226	226
Within R^2	.153	.174	.169
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

This table presents the effect of cultural distance between acquirers and targets on acquisition premiums. The dependent variable for all Models is the acquisition premium paid in relation to the target's market value, which is measured one month prior to the deal announcement. The main independent variable in Model 1 is *deal_culturaldistance*, with higher values indicating higher cultural dissimilarities between the acquirer and target. The main independent variables in Model 2 are the absolute cultural differences (in the four Competing Values Framework categories clan, adhocracy, market, hierarchy) between the acquirer and target. The main independent variables in Model 3 are binary variables indicating whether the market difference between the acquirer and target is ranked in the highest (lowest) quartile of all deals. A detailed variable description can be found in Appendix B. Constant terms are estimated but not reported. We include the acquirer's industry and year fixed effects in all models. We cluster standard errors at the 2-digit SIC industry level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Table 6
H3: Deal duration

	(1) deal_duration_ln	(2) deal_duration_ln	(3) deal_duration_ln
deal_culturaldistance	-.277		
	(-.8)		
deal_clan_culture_abs		-.431	
		(-.663)	
deal_adhocracy_culture_abs		-.25	
		(-.374)	
deal_market_culture_abs		.822	
		(1.452)	
deal_hierarchy_culture_abs		-1.11*	
		(-1.966)	
deal_hierarchy_diff_acq_high			-.201***
			(-2.716)
deal_hierarchy_diff_acq_low			-.024
			(-.255)
deal_value_ln	.161***	.166***	.152***
	(3.125)	(3.455)	(3.075)
deal_hofstede_distance	.654	.598	.538
	(1.062)	(.919)	(.883)
deal_all_cash_dummy	.025	.048	.026
	(.255)	(.499)	(.271)
deal_all_stock_dummy	-.003	-.007	-.005
	(-.022)	(-.048)	(-.036)
deal_tenderoffer_dummy	-.605***	-.614***	-.606***
	(-6.522)	(-6.686)	(-7.141)
deal_friendly_dummy	-.35	-.54*	-.434*
	(-1.204)	(-1.844)	(-1.999)
deal_relatedness_dummy	.111	.091	.095
	(.882)	(.846)	(.798)
deal_bidders	.124	.116	.129
	(.59)	(.587)	(.601)
acq_age	.001*	.001*	.001*
	(1.754)	(1.729)	(1.889)
acq_assets_lastyear_ln	-.077*	-.08**	-.075*
	(-1.95)	(-2.03)	(-1.919)
acq_roa_lastyear	-.05*	-.049	-.045*
	(-1.867)	(-1.523)	(-1.855)
acq_industrygrowth	-1.296	-1.614	-1.514*

	(-1.514)	(-1.45)	(-1.73)
acq_dealexperience	.003	.01	.011
	(.079)	(.243)	(.309)
tar_age	.001	.001	.001
	(1.173)	(1.038)	(.948)
tar_assets_lastyear_ln	.085	.075	.092
	(1.421)	(1.303)	(1.625)
tar_roa_lastyear	.251	.214	.224
	(.516)	(.472)	(.514)
tar_mbratio	.001	.003	.002
	(.263)	(.833)	(.85)
Observations	241	241	241
Adj R^2	.576	.582	.586
Industry FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

This table presents the effect of cultural distance between acquirers and targets on deal duration. The dependent variable for all Models is the natural logarithm of the number of days between deal announcement and deal completion. The main independent variable in Model 1 is *deal_culturaldistance*, with higher values indicating higher cultural dissimilarities between the acquirer and target. The main independent variables in Model 2 are the absolute cultural differences (in the four Competing Values Framework categories clan, adhocracy, market, hierarchy) between the acquirer and target. The main independent variables in Model 3 are binary variables indicating whether the hierarchy difference between the acquirer and target is ranked in the highest (lowest) quartile of all deals. A detailed variable description can be found in Appendix B. Constant terms are estimated but not reported. We include the acquirer's industry and year fixed effects in all models. We cluster standard errors at the 2-digit SIC industry level. t-statistics are reported in parentheses below the coefficients. *, **, *** denote statistical significance at the 0.10, 0.05, and 0.01 levels (using two-tailed tests).

Appendix

Appendix A – Glassdoor review examples associated with the four dimensions of the Competing Values Framework

Culture	Score	Sample Review Text
Adhocracy (create)	An adhocracy culture focuses on adaptability and flexibility to achieve growth and innovation within the organization.	
	0.0003	this review is for the office 365/outlook team.. * great people and management. * family friendly (great work-life balance). * great benefits. * great facilities (medical facilities, sports fields, i hear they even have a treehouse now). * free drinks cooler i've heard nightmares in certain teams, so ymmv depending on the team.. no free food.
	0.9956	fast paced, new products and technology, exciting opportunities and ability to try new and different things, support from management and colleagues constant re orgs . inconsistent messaging at times. travel. difficult/laborious to get someone promoted . too many systems and logins . holiday schedule
Clan (collaborate)	The clan culture emphasizes collaboration, teamwork, and employee development.	
	0.0003	discounts on services. decent pay. the company has changed a lot, they are only interested in pushing sales and not disclosing the proper information to the customer. providing good customer service is not a concern for them anymore. the information the call center gives customers is not the same as the stores. they are eliminating the need for full time employees. they are eliminating many jobs in the united states. sadly, you deal with a lot of angry and frustrated people. retail hours.
	0.9979	amazing work life balance, 1 on 1 sales training, friendly work environment, and opportunity to move up. the people here are very nice and the ages of everyone varies from mid twenties and up in a balanced matter. everyone here wants to succeed and that energy is passed on to all employees. The only thing i wish we had were nicer bathrooms, but i can deal with that! commission structure could be better as well. not the best, but modest.
Market (compete)	A market culture tries to maximize business or production performance by focusing on task completion and goal achievement.	
	0.0003	culture management good learning compensation and benefits policies flexibility cafe vaccination drive well equipped gym inhouse doctors, nurses, nutrition, gym coach and clinic nothing major i can think of. enjoyed working in the company and a great place to learn. inter-department teams work together.
	0.9970	nothing is worth the stress and aggregation they put you through. stay away if you can. salary is competitive. 401k is ok. large company so easy to stay close to home

		if they permit. Overworked as if they are legally breaking labor laws. horrible management too down. management pushes you to fake numbers to improve metrics. questionable patient safety practices in pharmacy. hazardous work conditions many times on the sale floor as there is not enough hours/work ratio to finish work.
Hierarchy (control)	The hierarchical culture emphasizes clear rules, explicit instructions, and strict controls.	
	0.0005	the managers are great people, and very kind. i love all of my coworkers, and i love the work. copy center is fast paced and always different. love getting to know the products and learning as i went. the company makes cuts in the wrong places. cutting part time hours to under 25 a week, to save $4 million a year. but sending the higher ups on vacations. not enough hours. obviously, no benefits.
	0.9931	great benefits for full time employees. it's corporate retail, so long periods of standing, and micro managing everything you do. but the biggest problem is stagnant wages, and when you do get a yearly raise it's in the 1-3% range. not somewhere for a career, unless you want to give up most of your personal time and become a salary slave. then your still going to get small raises, your able to compensate somewhat with the store bonus, depending on your store sales.

Score represents the probability score inferred from the CultureBERT transformer model, which was manually pre-trained on 2,000 Glassdoor reviews (Koch and Pasch 2022). Higher values indicate a high affiliation to the respective Competing Values Framework (Cameron et al. 2006) culture dimension. Since the dimensions are not mutually exclusive, a firm's review can have high affiliations with multiple dimensions.

Appendix B – Variable definitions

Variable	Definition
Independent variables	
acq_clan_culture	Probability score is determined by applying the CultureBERT transformer (Koch & Pasch, 2022) to firms' Glassdoor textual reviews. Bound between 0 and 1, with values close to 1 indicating a high clan affiliation. Firms' culture scores are first averaged by year and then averaged over all years until the deal announcement. Analogously for the other three competing value framework dimensions (adhocracy, market, hierarchy). Examples of the four dimensions are provided in Appendix A
acq_clan_dominant	Binary variable indicating whether the clan dimension has the highest probability score of all four culture dimensions. Analogously for the other three competing value framework dimensions (adhocracy, market, hierarchy)
deal_clan_culture_abs	Absolute difference between the acquirer's and target's clan culture. Analogously for the other three competing value framework dimensions (adhocracy, market, hierarchy)
deal_clan_diff_acq_high	Binary variable indicating whether the clan difference between the acquirer and target is ranked in the highest quartile of all deals. Analogously for the other three competing value framework dimensions (adhocracy, market, hierarchy)
deal_culturaldistance	Euclidean distance between the acquirer's and target's four cultural Competing Values Framework dimensions. Higher values indicate a higher cultural distance between the acquirer and the target. As a robustness check, the variable is also calculated using the Jensen-Shannon divergence
Dependent variables	
acq_car2, acq_car5	Acquirer's cumulative abnormal returns estimated using the market model over the period of 50 to 250 days prior to the announcement date
deal_car_weighted	Value-weighted combination of the acquirer's and target's cumulative abnormal returns [-5, 5] days around the announcement day, using the relative market values as weights

deal_premium_1month	Difference between acquirer's payment and target's market value, divided by target's market value, measured one month (one day; one week) before the deal announcement
acq_ros_1year, acq_ros_3years	Change in return on sales (operating profit divided by net sales) one year (three years) after the deal announcement in relation to the firm's return on sales during the deal announcement year.
deal_duration_ln	Natural logarithm of the difference in days between the date the deal was completed and the date when the deal was announced for the first time by an involved party
Deal controls	
deal_value_ln	Natural logarithm of the total value of consideration paid by the acquirer, excluding fees and reported expenses
deal_all_cash_dummy	Dummy variable that equals one if the deal was fully paid in cash
deal_all_stock_dummy	Dummy variable that equals one if the deal was fully paid in stocks
deal_tenderoffer_dummy	Dummy variable that equals one when a tender offer is launched for the target
deal_friendly_dummy	Dummy variable that equals one if the deal is marked as friendly
deal_relatedness	Dummy variable that equals one if the acquirer and the target operate in the same two-digit SIC code
deal_bidders	Number of competing bidders
deal_hofstede_distance	National cultural distance between the acquirer's and target's nations, computed as the Euclidean distance of Hofstede's six cultural dimensions (Individualism, Power Distance, Uncertainty Avoidance, Femininity, Indulgence, Long-term orientation). Each distance is bound between 0 and 1
Acquirer/target controls (*acq_ / tar_ prefixes*)	
firm_age	Difference between the year when the transaction was completed and the year when the firm was founded
firm_assets_lastyear_ln	Logarithm of the total assets of the firm in the last 12 months before the deal announcement
firm_roa_lastyear	Ratio of the firm's net income to total assets, measured 12 months before the deal announcement

firm_mbratio	Ratio of firm’s market capitalization to book value of total assets at the end of the fiscal year prior to deal announcement
acq_industrygrowth	Average percentage change in revenue for the acquirer’s 2-digit SIC industry sector, divided by the revenue reported in the year prior to deal announcement
acq_dealexperience	Number of deals successfully completed by the acquirer in the last three years prior to the announcement date, including the current deal

References

Ahern, K. R., Daminelli, D., & Fracassi, C. (2015). Lost in translation? The effect of cultural values on mergers around the world. *Journal of Financial Economics, 117*, 165–189.

Ahmed, A. S., Chen, L. H., Duellman, S., & Sun, Y. (2023). Targets' Accounting Conservatism and the Gains from Acquisition. *Contemporary Accounting Research, 40*, 7–40.

Akerlof, G. A. (1997). Social Distance and Social Decisions. *Econometrica, 65*, 1005.

Aktas, N., Bodt, E. de, & Cousin, J.-G. (2011). Do financial markets care about SRI? Evidence from mergers and acquisitions. *Journal of Banking & Finance, 35*, 1753–1761.

Alexandridis, G., Hoepner, A. G., Huang, Z., & Oikonomou, I. (2022). Corporate social responsibility culture and international M&As. *The British Accounting Review, 54*, 101035.

Arrow, K. J. (1974). *The Limits of Organization*. New York: W.W. Norton.

Aybar, B., & Ficici, A. (2009). Cross-border acquisitions and firm value: An analysis of emerging-market multinationals. *Journal of International Business Studies, 40*, 1317–1338.

Barkema, H., Bell, J., & Pennings, J. (1996). Foreign Entry, Cultural Barriers, and Learning. *Strategic Management Journal, 17*(2), 151–166.

Barraclough, K., Robinson, D. T., Smith, T., & Whaley, R. E. (2013). Using Option Prices to Infer Overpayments and Synergies in M&A Transactions. *The Review of Financial Studies, 26*, 695–722.

Bena, J., & Li, K. (2014). Corporate Innovations and Mergers and Acquisitions. *The Journal of Finance, 69*, 1923–1960.

Bereskin, F., Byun, S. K., Officer, M. S., & Oh, J.-M. (2018). The Effect of Cultural Similarity on Mergers and Acquisitions: Evidence from Corporate Social Responsibility. *Journal of Financial and Quantitative Analysis, 53*, 1995–2039.

Bhagat, R. S., & McQuaid, S. J. (1982). Role of subjective culture in organizations: A review and directions for future research. *The Journal of applied psychology, 67*, 653–685.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2001). Latent Dirichlet Allocation. *Advances in Neural Information Processing Systems 14, 3*(Jan), 601–608.

Bochkay, K., Brown, S. V., Leone, A. J., & Tucker, J. W. (2023). Textual Analysis in Accounting: What's Next?. *Contemporary Accounting Research, 40*, 765–805.

Bouwman, C. H. S. (2013). *The Role of Corporate Culture in Mergers & Acquisitions*.

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of Financial Economics, 14*, 3–31.

Buono, A. F., Bowditch, J. L., & Lewis, J. W. (1985). When Cultures Collide: The Anatomy of a Merger. *Human Relations, 38*, 477–500.

Büschgens, T., Bausch, A., & Balkin, D. B. (2013). Organizational Culture and Innovation: A Meta-Analytic Review. *Journal of Product Innovation Management, 30*, 763–781.

Cameron, K., Quinn, R., DeGraff, J., & Thakor, A. (2006). *Competing values leadership: Creating value in organizations*. Cheltenham, U.K, Northampton, Mass: Edward Elgar.

Campbell, D. W., & Shang, R. (2021). Tone at the Bottom: Measuring Corporate Misconduct Risk from the Text of Employee Reviews. *Management Science*.

Cartwright, S., & Cooper, C. L. (1993). The role of culture compatibility in successful organizational marriage. *Academy of Management Perspectives, 7*, 57–70.

Chakrabarti, R., Gupta-Mukherjee, S., & Jayaraman, N. (2009). Mars–Venus marriages: Culture and cross-border M&A. *Journal of International Business Studies, 40*, 216–236.

Chatman, J. A., & O'Reilly, C. A. (2016). Paradigm lost: Reinvigorating the study of organizational culture. *Research in Organizational Behavior, 36*, 199–224.

Chatterjee, A., & Hambrick, D. C. (2011). Executive Personality, Capability Cues, and Risk Taking. *Administrative Science Quarterly, 56*, 202–237.

Chatterjee, S., Lubatkin, M. H., Schweiger, D. M., & Weber, Y. (1992). Cultural differences and shareholder value in related mergers: Linking equity and human capital. *Strategic Management Journal, 13*, 319–334.

Chemmanur, T. J., Rajaiya, H., & Sheng, J. (2019). How Does Soft Information Affect External Firm Financing? Evidence from Online Employee Ratings.

Chen, C.-W., Collins, D. W., Kravet, T. D., & Mergenthaler, R. D. (2018). Financial Statement Comparability and the Efficiency of Acquisition Decisions. *Contemporary Accounting Research, 35*, 164–202.

Chen, J., & Liu, L. (2022). Family firms, national culture and corporate social performance: a meta-analysis. *Cross Cultural & Strategic Management, 29*, 379–402.

Cho, S., & Chung, C. Y. (2022). Review of the Literature on Merger Waves. *Journal of Risk and Financial Management, 15*, 432.

Conn, R. L., Cosh, A., Guest, P. M., & Hughes, A. (2005). The Impact on UK Acquirers of Domestic, Cross-border, Public and Private Acquisitions. *Journal of Business Finance & Accounting, 32*, 815–870.

Corritore, M., Goldberg, A., & Srivastava, S. B. (2020). Duality in Diversity: How Intrapersonal and Interpersonal Cultural Heterogeneity Relate to Firm Performance. *Administrative Science Quarterly, 65*, 359–394.

Dahlgrün, P. W., & Bausch, A. (2019). How Opportunistic Culture Affects Financial Performance in Outsourcing Relationships: A Meta-Analysis. *Journal of International Management, 25*, 81–100.

Datta, D. K., & Puia, G. (1995). Cross-Border Acquisitions: An Examination of the Influence of Relatedness and Cultural Fit on Shareholder Value Creation in U.S. Acquiring Firms. *Management International Review, 35*(4), 337–359.

Dikova, D., Sahib, P. R., & van Witteloostuijn, A. (2010). Cross-Border Acquisition Abandonment and Completion: The Effect of Institutional Differences and Organizational Learning in the International Business Service Industry. *Journal of International Business Studies, 41*(2), 223–245.

Ellis, K. M., Reus, T. H., Lamont, B. T., & Ranft, A. L. (2011). Transfer Effects in Large Acquisitions: How Size-Specific Experience Matters. *Academy of Management Journal, 54*, 1261–1276.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance, 25*, 383.

Ghosh, A., & Ruland, W. (1998). Managerial Ownership, the Method of Payment for Acquisitions, and Executive Job Retention. *The Journal of Finance, 53*, 785–798.

Giliberto, M., & Varaiya, N. (1989). The Winner's Curse and Bidder Competition in Acquisitions: Evidence from Failed Bank Auctions. *The Journal of Finance, 44*, 59–75.

Haspeslagh, P. C., & Jemison, D. B. (1991). The challenge of renewal through acquisitions. *Planning Review, 19*, 27–30.

Haunschild, P. R. (1994). How Much is That Company Worth?: Interorganizational Relationships, Uncertainty, and Acquisition Premiums. *Administrative Science Quarterly, 39*, 391.

Hayward, M. L. A., & Hambrick, D. C. (1997). Explaining the Premiums Paid for Large Acquisitions: Evidence of CEO Hubris. *Administrative Science Quarterly, 42*, 103.

Hofstede, G. (1980). Culture and Organizations. *International Studies of Management & Organization, 10*, 15–41.

Hofstede, G. H. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd ed.). Thousand Oaks, Calif., London: SAGE.

Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why Differences Make a Difference: A Field Study of Diversity, Conflict and Performance in Workgroups. *Administrative Science Quarterly, 44*, 741–763.

Jentner, D., & Lewellen, K. (2015). CEO Preferences and Acquisitions. *The Journal of Finance, 70*, 2813–2852.

King, D. R., Wang, G., Samimi, M., & Cortes, A. F. (2021). A Meta-Analytic Integration of Acquisition Performance Prediction. *Journal of Management Studies, 58*, 1198–1236.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A quarter century of Culture's Consequences: a review of empirical research incorporating Hofstede's cultural values framework. *Journal of International Business Studies, 37*, 285–320.

Koch, S., & Pasch, S. (2022). CultureBERT: Fine-Tuning Transformer-Based Language Models for Corporate Culture.

Kogut, B., & Singh, H. (1988). The Effect of National Culture on the Choice of Entry Mode. *Journal of International Business Studies, 19*, 411–432.

Laamanen, T. (2007). On the role of acquisition premium in acquisition research. *Strategic Management Journal, 28*, 1359–1369.

Lawrence, E. R., Raithatha, M., & Rodriguez, I. (2021). The effect of cultural and institutional factors on initiation, completion, and duration of cross-border acquisitions. *Journal of Corporate Finance, 68*, 101950.

Lee, G., Cho, S. Y., Arthurs, J., & Lee, E. K. (2019). CEO pay inequity, CEO-TMT pay gap, and acquisition premiums. *Journal of Business Research, 98*, 105–116.

Lee, K. H. (2018). Cross-border mergers and acquisitions amid political uncertainty: A bargaining perspective. *Strategic Management Journal, 39*, 2992–3005.

Li, C., & Haleblian, J. (2022). The Influence of Nation-Level Institutions on Acquisition Premiums: A Cross-Country Comparative Study. *Journal of Management, 48*, 878–904.

Li, K., Mai, F., Shen, R., & Yan, X. (2021). Measuring Corporate Culture Using Machine Learning. *The Review of Financial Studies, 34*, 3265–3315.

Lim, J., Makhija, A. K., & Shenkar, O. (2016). The asymmetric relationship between national cultural distance and target premiums in cross-border M&A. *Journal of Corporate Finance, 41*, 542–571.

Liu, Y., Ott, M., Goyal, N., Du Jingfei, Joshi, M., Chen, D., et al. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. https://arxiv.org/pdf/1907.11692.

Lorsch, J. W. (1986). Managing Culture: The Invisible Barrier to Strategic Change. *California Management Review, 28*, 95–109.

Loughran, T., & Vijh, A. M. (1997). Do Long-Term Shareholders Benefit From Corporate Acquisitions? *The Journal of Finance, 52*, 1765.

Louis, M. R. (1981). A cultural perspective on organizations: the need for and consequences of viewing organizations as culture-bearing milieux. *Human Systems Management, 2*, 246–258.

Martin, J. (1992). *Cultures in Organizations–Three Perspectives*. Oxford University Press.

McNichols, M. F., & Stubben, S. R. (2015). The effect of target-firm accounting quality on valuation in acquisitions. *Review of Accounting Studies, 20*, 110–140.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient Estimation of Word Representations in Vector Space*. https://arxiv.org/pdf/1301.3781.

Moeller, S. B., Schlingemann, F. P., & Stulz, R. M. (2004). Firm size and the gains from acquisitions. *Journal of Financial Economics, 73*, 201–228.

Morck, R., Shleifer, A., & Vishny, R. W. (1990). Do Managerial Objectives Drive Bad Acquisitions? *The Journal of Finance, 45*, 31.

Morosini, P., Shane, S., & Singh, H. (1998). National Cultural Distance and Cross-Border Acquisition Performance. *Journal of International Business Studies, 29*, 137–158.

Mosel, J. von der, Trautsch, A., & Herbold, S. (2023). On the Validity of Pre-Trained Transformers for Natural Language Processing in the Software Engineering Domain. *IEEE Transactions on Software Engineering, 49*, 1487–1507.

Naranjo-Valencia, J. C., Jiménez-Jiménez, D., & Sanz-Valle, R. (2011). Innovation or imitation? The role of organizational culture. *Management Decision, 49*, 55–72.

Nguyen, N. H., & Phan, H. V. (2017). Policy Uncertainty and Mergers and Acquisitions. *Journal of Financial and Quantitative Analysis, 52*, 613–644.

Oberoi, S. (2020). 6 Examples of Merger Failures Owing to Cultural Incompatibility. https://www.spiceworks.com/hr/hr-strategy/articles/examples-merger-failure-cultural-incompatibility/.

Qiao, L., & Wu, J. (2019). Pay for Being Responsible: The Effect of Target Firm's Corporate Social Responsibility on Cross-Border Acquisition Premiums. *Sustainability, 11*, 1291.

Quinn, R. E., & Rohrbaugh, J. (1983). A Spatial Model of Effectiveness Criteria: Towards a Competing Values Approach to Organizational Analysis. *Management Science, 29*, 363–377.

Raghavendra, R., & Vermaelen, T. (1998). Glamour, value and the post-acquisition performance of acquiring firms. *Journal of Financial Economics, 49*, 223–253.

Rahahleh, N. A., & Wei, P. P. (2013). Frequent cross-border acquirers from emerging countries and cultural distance: Does the cultural difference of the initial deal matter? *Journal of Multinational Financial Management, 23*, 356–373.

Recendes, T., Aime, F., Hill, A. D., & Petrenko, O. V. (2022). Bargaining your way to success: The effect of Machiavellian chief executive officers on firm costs. *Strategic Management Journal, 43*, 2012–2041.

Renneboog, L., & Vansteenkiste, C. (2019). Failure and success in mergers and acquisitions. *Journal of Corporate Finance, 58*, 650–699.

Reuer, J. J., Tong, T. W., & Wu, C.-W. (2012). A Signaling Theory of Acquisition Premiums: Evidence from IPO Targets. *Academy of Management Journal, 55*, 667–683.

Roll, R. (1986). The Hubris Hypothesis of Corporate Takeovers. *The Journal of Business, 59*, 197.

Rossi, S., & Volpin, P. F. (2004). Cross-country determinants of mergers and acquisitions. *Journal of Financial Economics, 74*, 277–304.

Rottig, D. (2017). Meta-analyses of culture's consequences for acquisition performance. *International Journal of Emerging Markets, 12*, 8–37.

Sarala, R. M., & Vaara, E. (2010). Cultural differences, convergence, and crossvergence as explanations of knowledge transfer in international acquisitions. *Journal of International Business Studies, 41*, 1365–1390.

Schein, E. H. (1985). *Organizational Culture and Leadership*.

Schweiger, D. M. (2002). *M&A Integration—A Framework for Executives & Manager*.

Schwert, G. W. (2000). Hostility in Takeovers: In the Eyes of the Beholder? *The Journal of Finance, 55*(6), 2599–2640.

Sirower, M. L. (1997). *The Synergy Trap: How Companies Lose the Acquisition Game*.

Sitkin, S. B., & Stickel, D. (1996). The Road to Hell: The Dynamics of Distrust in an Era of Quality. In R. Kramer & T. Tyler (Eds.), *Trust in Organizations: Frontiers of Theory and Research* (pp. 196–215). 2455 Teller Road, Thousand Oaks California 91320 United States: SAGE Publications, Inc.

Sørensen, J. B. (2002). The Strength of Corporate Culture and the Reliability of Firm Performance. *Administrative Science Quarterly, 47*, 70–91.

Stahl, G. K., & Voigt, A. (2008). Do Cultural Differences Matter in Mergers and Acquisitions? A Tentative Model and Examination. *Organization Science, 19*, 160–176.

Suk, I., & Wang, M. (2021). Does target firm insider trading signal the target's synergy potential in mergers and acquisitions? *Journal of Financial Economics, 142*(3), 1155–1185.

Tajfel, H. (1981). *Human Groups and Social Categories—Studies in Social Psychology*. Cambridge: Cambridge University Press.

Teerikangas, S., & Very, P. (2006). The Culture-Performance Relationship in M&A: From Yes/No to How. *British Journal of Management, 17*, S31-S48.

Turner, J. (1982). *Toward a cognitive definition of the group. In H. Tajfel (Ed.), Social Identity and Intergroup Relations*. Cambridge: Cambridge University Press.

Vaara, E. (2002). On the Discursive Construction of Success/Failure in Narratives of Post-Merger Integration. *Organization Studies, 23*, 211–248.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). *Attention Is All You Need*. https://arxiv.org/pdf/1706.03762.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural Diversity's Impact on Interaction Process and Performance: Comparing Homogeneous and Diverse Task Groups. *Academy of Management Journal, 36*, 590–602.

Weber, R. A., & Camerer, C. F. (2003). Cultural Conflict and Merger Failure: An Experimental Approach. *Management Science, 49*, 400–415.

Weber, Y. (1996). Corporate Cultural Fit and Performance in Mergers and Acquisitions. *Human Relations, 49*, 1181–1202.

Zhu, H., Ma, X., Sauerwald, S., & Peng, M. W. (2019). Home Country Institutions Behind Cross-Border Acquisition Performance. *Journal of Management, 45*, 1315–1342.

Zollo, M., & Meier, D. (2008). What Is M&A Performance? *Academy of Management Perspectives, 22*, 55–77.

Zollo, M., & Singh, H. (2004). Deliberate learning in corporate acquisitions: post-acquisition strategies and integration capability in U.S. bank mergers. *Strategic Management Journal, 25*, 1233–1256.

Machine Learning vs. Management Forecasts:
Können Machine-Learning-Modelle die Genauigkeit von Umsatzprognosen verbessern?

Autoren Hannes Gerstel, Mohamed Khaled

Zusammenfassung Möglichst akkurate Umsatzprognosen sind relevant für Unternehmen, um Unsicherheiten im Planungs- und Budgetierungsprozess zu minimieren. Der Beitrag zeigt, dass die Genauigkeit von Umsatzprognosen durch Machine-Learning-basierte Forecasts im Vergleich zu Management Forecasts um etwa 20 % verbessert werden kann. Neben der erhöhten Prognosegenauigkeit ergeben sich potenzielle Zeit- und Kostenersparnisse.

Publikationsstatus Veröffentlicht (https://doi.org/10.15358/0935-0381-2023-4)
Controlling – Zeitschrift für erfolgsorientierte Unternehmenssteuerung
2023 Heft 4 (VHB-Jourqual 3: D)

Danksagung Wir danken Sophie Kulig für ihre hilfreichen Anmerkungen.

Author Contribution	Gerstel	Khaled
Relative Share	*95%*	*5%*
Conceptualization	✓	
Theory Development	✓	
Methodology	✓	✓
Data Acquisition	✓	
Data Analyses	✓	
Writing	✓	

1. Die Relevanz akkurater Umsatzprognosen

Im Jahr 2022 haben etwa 1.000 börsengelistete US-Unternehmen vierteljährliche Umsatzprognosen im Rahmen ihrer Earnings Calls freiwillig veröffentlicht (Earnings Guidance). Damit Stakeholder (bspw. Kapitalmarktteilnehmer als auch das Unternehmen selbst) sich auf die Forecasts verlassen können, ist ein möglichst geringer Prognosefehler Voraussetzung. Akkurate Forecasts helfen dem Management insbesondere dabei, Unsicherheiten im Planungs- und Budgetierungsprozess zu reduzieren (vgl. *Goretzki/Wiegmann*, 2022). Zum Beispiel können Prognosen signalisieren, dass geplante Unternehmensziele voraussichtlich verfehlt werden, wodurch möglichst frühzeitig Gegenmaßnahmen eingeleitet werden können.

Dieser Beitrag soll die Frage beantworten, ob die Genauigkeit von Umsatzprognosen durch Anwendung von Machine Learning (ML)-Modellen verbessert werden kann. ML beschreibt ein Teilgebiet der künstlichen Intelligenz (vgl. Hastie et al. 2009), in dem Muster in Daten mittels verschiedener Algorithmen (z.B. lineare Regression oder Decision Trees) identifiziert werden sollen. Unsere Ergebnisse zeigen, dass der Prognosefehler von vierteljährlichen Umsatzschätzungen des Managements mittels ML-Modellen im Durchschnitt um 20 % reduziert werden kann. Zum Beispiel hat *Visa Inc.* im Earnings Call des dritten Fiskalquartals 2022 einen Umsatz von 7.019 Mio. $ prognostiziert. Drei Monate später hat *Visa Inc.* bekanntgegeben, dass ein Umsatz von 7.275 Mio. $ realisiert wurde. Dadurch ergibt sich ein Schätzfehler von 3,5 %. Ein von uns mit Vergangenheitsdaten trainiertes ML-Modell liefert in diesem Beispiel einen Prognosewert von 7.377 Mio. $, wodurch sich der Prognosefehler auf 1,4 % reduziert.

Da Earnings Guidances –im Gegensatz zu unternehmensinternen Forecasts– öffentlich verfügbar sind, nutzen wir diese in unseren Analysen. Neben der Relevanz der Prognosegenauigkeit für den Kapitalmarkt existieren nachweisbare Zusammenhänge zwischen der Qualität der Earnings Guidance und dem unternehmensinternen Forecasting. So ist oftmals die Vorstandsvergütung an die Prognosegenauigkeit der Earnings Guidance gekoppelt (vgl. *Hui/Matsunaga*, 2015) und die Verwendung von hochwertigen internen Planungs- und Forecasting-Prozessen hängt positiv mit der Prognosegenauigkeit der Earnings Guidance zusammen (vgl. *Ittner/Michels*, 2017). Außerdem existiert ein positiver Zusammenhang zwischen akkurater Earnings Guidance und dem Erfolg von Investitionen, welche auf der internen Budgetierung und Prognosen basieren (vgl. *Goodman* et al.*,* 2014). Des Weiteren nutzen einige Unternehmen bereits erfolgreich ML-Komponenten in deren internen Forecasts. Durch Komplementierung der klassischen, menschlichen Forecasts durch ML-Modelle konnte bspw. *Microsoft* den durchschnittlichen Prognosefehler der internen Forecasts von ca. 2,7 % auf 1,5 % senken (vgl. *Wiprächtiger*, 2021). Folglich ist davon auszugehen, dass die Anwendung von ML nicht nur Earnings Guidances verbessern kann, sondern auch das interne Forecasting von ML profitieren kann.

Im Folgenden werden zunächst Grundlagen zu den eingesetzten ML-Modellen dargestellt. Anschließend werden das empirische Vorgehen und die Ergebnisse vorgestellt, um anschließend Implikationen für die Praxis zu diskutieren.

2. Grundlagen zur Anwendung von Machine-Learning-Modellen

Relevante Problemfelder von Machine Learning

Im ML wird zwischen unsupervised und supervised Problemen differenziert (vgl. Hastie et al. 2009). Bei unsupervised Problemen sollen Muster zwischen Beobachtungen ohne zuvor bekannte Zielvariable identifiziert werden, bspw. durch Clustering. Beim supervised ML sollen mittels erklärender Daten („Features“) eine Zielvariable (bspw. zukünftiger Umsatz) vorhergesagt werden. Als Features werden bspw. der vergangene Umsatz und die Marktkapitalisierung als Input für die Algorithmen verwendet. Des Weiteren wird der Datensatz generell in einen Trainings- und einen Testdatensatz unterteilt. Mit den Trainingsdaten wird das jeweilige Modell trainiert, um dann anschließend Prognosen der Zielvariable für den Testdatensatz zu generieren. Bei der durchgeführten Analyse handelt es sich um ein Regressionsproblem, da wir kontinuierliche Umsätze als Zielvariable verwenden.

Die Modellauswahl für das vorliegende Regressionsproblem basiert auf hochrangiger, aktueller Forschung. *Chen* et al. (2022) finden heraus, dass Random Forests und Gradient Boosted Trees das Vorzeichen des nächsten einjährigen Gewinnwachstums im Vergleich zu Analystenprognosen akkurater vorhersagen können. Allerdings werden deren Modelle von Analysten geschlagen, sobald absolute Umsatzprognosen betrachtet werden. Ding et al. (2020) zeigen, dass mittels Random Forest geschätzte Schadensrückstellung in der Versicherungsbranche größtenteils zu genaueren Prognosen kommen als die Schätzungen des Managements, welche verpflichtend in den Jahresabschlüssen angegeben werden müssen. Neben der Forschung orientieren wir uns an den erfolgreichsten ML-Modellen für tabellarische Regressionsprobleme der ML- und Data-Science-Webseite *kaggle.com*. Dort suchen Unternehmen die präzisesten Modelle für ihre ML-Probleme. Die Lösungen dazu sind öffentlich einsehbar und mit bis zu 7-stelligen Preisgeldern dotiert. Bei einem Wettbewerb zu Umsatzprognosen in *Walmart*-Filialen lieferten Gradient Boosted Trees die besten Ergebnisse (vgl. Makridakis et al. 2022). Basierend auf diesen Ergebnissen nutzen wir zur Erstellung der Umsatzprognosen ebenfalls (optimierte) Decision Trees.

Decision Trees

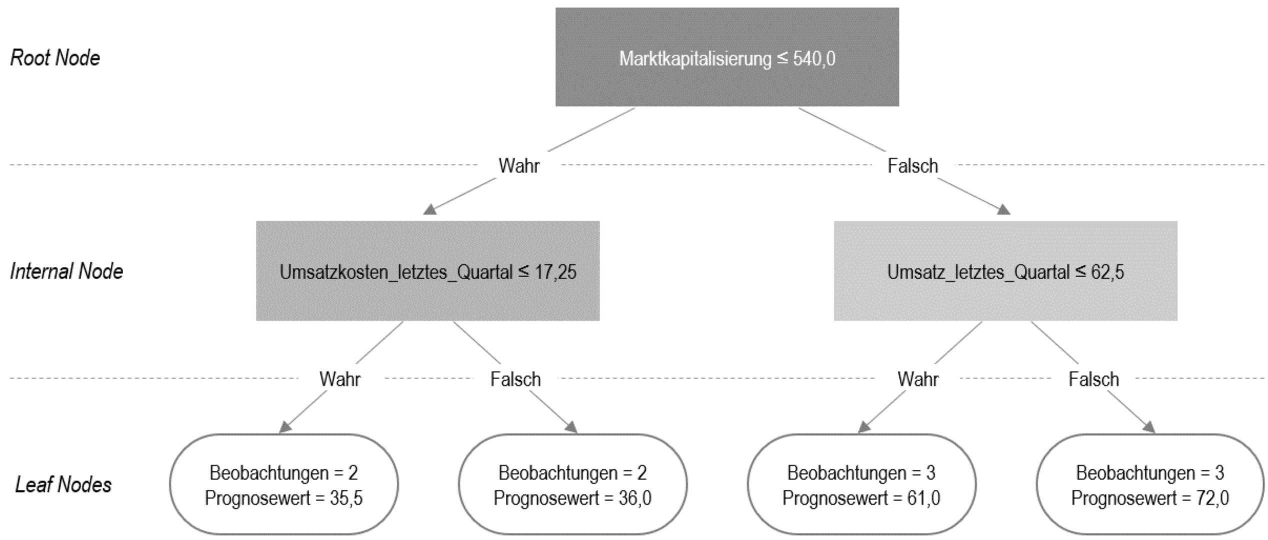


Abb. 1: Illustratives Decision-Tree-Modell

Insbesondere im Vergleich zu populären linearen Modellen (wie bspw. der OLS-Regression) besteht bei Decision Trees der Vorteil, dass nicht-lineare Zusammenhänge und Interaktionen zwischen den Features modelliert werden können (vgl. Hastie et al. 2009). *Abb. 1* erklärt den Algorithmus eines Decision Trees. Dazu werden 10 illustrative Unternehmensbeobachtungen als Datensatz verwendet, in denen der aktuelle Quartalsumsatz die Zielvariable darstellt. Als Features werden die Marktkapitalisierung, der vorherige Quartalsumsatz und die Umsatzkosten genutzt. Das Ziel des Decision Trees ist es, jeder Beobachtung einen geschätzten Umsatz mit möglichst geringem Prognosefehler zuzuordnen, indem binäre Teilungen der Features erfolgen. Der Startpunkt des Baumes (Root Node) wird bestimmt, indem das Feature Marktkapitalisierung genau an dem Wert geteilt wird, an welchem der Prognosefehler minimal ist. In dem fiktiven Beispiel in *Abb. 1* werden Unternehmen mit einer Marktkapitalisierung von kleiner gleich und größer 540 Mio. $ in je eine Gruppe eingeteilt. Theoretisch hätte auch ein anderes Feature an der Root Node stehen können, allerdings ist die Minimierung des Fehlerterms bei den anderen Features geringer. Da die Anzahl der maximalen Ebenen des Baums in diesem Beispiel auf zwei beschränkt wird, evaluiert der Algorithmus erneut, ob anhand der Features weitere sinnvolle Teilungen vorgenommen werden können. Dadurch entstehen zwei weitere Teilungen

(Internal Nodes). Die linke Node teilt Beobachtungen mit kleiner Marktkapitalisierung erneut in zwei Gruppen, falls diese Umsatzkosten von kleiner gleich 17,25 Mio. $ vorweisen. Zwei Beobachtungen unterschreiten die zwei Grenzwerte der Marktkapitalisierung und der Umsatzkosten. Der Mittelwert des aktuellen Umsatzes dieser zwei Beobachtungen liegt bei 35,5 Mio. $. Der Wert dieser „Leaf Node" (das Endstück links unten in *Abb. 1*) entspricht dem geschätzten Umsatz für die zwei Unternehmen. Die Splits und Prognosewerte aus dem mit Trainingsdaten angelernten Modell werden anschließend genutzt, um die Prognosen für die Beobachtungen aus dem Testdatensatz zu erstellen.

Random Forest und Gradient Boosted Trees

Durch Optimierung der Hyperparameter, bspw. der Tiefe des Baums, kann der Fehlerterm meistens stark reduziert werden. Dabei entsteht allerdings das Problem, dass die antrainierte Baumstruktur vermehrt ungenaue Ergebnisse für bisher unbekannte Beobachtungen liefert. Um diesem „Overfitting" entgegenzuwirken, werden oftmals mehrere Decision Trees kombiniert. Beim **Random-Forest**-Verfahren kommt „Bagging" zum Einsatz, bei dem mehrere einzelne Decision Trees auf zufällig ausgewählte Teilbeobachtungen und Features modelliert werden. Um den finalen Schätzwert des Random Forest für eine Beobachtung zu bestimmen, wird ein Durchschnittswert der einzelnen Schätzwerte der Decision Trees gebildet.

Eine weitere Verbesserung der Prognosegenauigkeit kann mit **Gradient Boosted Trees** erzielt werden, die auf aufeinander aufbauenden Bäumen basieren. Dabei werden die jeweils geschätzte Prognosefehler der weiteren Bäume mit dem Learning-Rate-Parameter skaliert und iterativ auf die initiale Schätzung des ersten Baums addiert („Boosting"). Dieser Schritt wird so lange wiederholt, bis sich die Modellgenauigkeit nicht mehr verbessert oder das vorher definierte Maximum an Iterationen erreicht ist. Unsere Analysen basieren größtenteils auf dem von *Microsoft* entwickelten *lightGBM* Algorithmus (vgl. Ke et al. 2017). Dieser zählt oftmals zu den erfolgreichsten und effizientesten Modellen bei *Kaggle*-Wettbewerben (vgl. Makridakis et al. 2022).

3. Empirisches Machine-Learning-Modell

Stichprobe

Für unsere Analyse nutzen wir 1.661 börsennotierte US-Firmen aus der *S&P CapitalIQ*-Plattform, die eine Marktkapitalisierung von mindestens 10 Mio. $ haben. Da Umsatzprognosen erst ab 2001 flächendeckend in *CapitalIQ* verfügbar sind, bezieht sich unser Untersuchungszeitraum auf die Jahre 2001 bis 2022. Aufgrund der Anfälligkeit von Decision-Tree-Verfahren gegenüber Ausreißern in der Zielvariable (vgl. Geertsema and Lu 2023), entfernen wir die 10 % umsatzstärksten Beobachtungen. Die Prognoseergebnisse dieser Beobachtungen werden in einem separaten Modell in Kapitel 4 gezeigt. Aus Plausibilitätsgründen entfernen wir Beobachtungen mit negativen Umsätzen. Fehlende Werte bei nicht-essenziellen Features (die keinen direkten Bezug zum Umsatz haben) werden durch null ersetzt. Beobachtungen mit fehlender Zielvariable werden vollständig entfernt. Der finale Datensatz besteht aus 19.998 Quartalsbeobachtungen mit einem durchschnittlichen Umsatz von 329 Mio. $.

Zielvariable und Features

Die **Zielvariable** entspricht den Quartalsumsätzen der Unternehmen. Die Umsatzprognosen des Managements werden entweder als Bandbreite oder als Punktschätzung angegeben. Falls eine Bandbreite angegeben ist, dann wird auf den Mittelpunkt der Bandbreite zurückgegriffen. Als Evaluationsmetrik benutzen wir für die Management- und die ML-Schätzungen den Mean Absolute Error (MAE = Prognosefehler) (vgl. *Ding* et al., 2020):

$$MAE\ (Prognosefehler) = \frac{\sum_{i=1}^{n} |y_{i,t} - \hat{y}_{i,t}|}{n}$$

n entspricht der Zahl an Quartalsbeobachtungen aller betrachteter Unternehmen über den Untersuchungszeitraum. Je Quartalsbeobachtung t eines Unternehmens i wird der absolute Prognosefehler berechnet (damit sich Über- und Unterschätzungen nicht ausgleichen), indem der geschätzte Quartalsumsatz ($\hat{y}_{i,t}$) vom tatsächlichen Quartalsumsatz ($y_{i,t}$) subtrahiert wird. Von den einzelnen Prognosefehlern in einem Zeitraum wird anschließend der Durchschnitt berechnet. Der Prognosevorteil wird ermittelt, indem der MAE der ML-Modelle und der Manager in Relation gesetzt wird (vgl. *Ding* et al., 2020):

$$Prognosevorteil = \frac{MAE_{Manager} - MAE_{ML}}{MAE_{Manager}}$$

Unser finales Modell nutzt **25 Features**, die auf Bilanz- und GuV-Informationen (bspw. Anlagevermögen und Umsatzkosten), markt- und makroökonomischen Informationen (bspw. Aktienkursänderungen und Bruttoinlandsprodukt) sowie den Management-Umsatzprognosen basieren. Das Modell ist dabei ausschließlich mit Daten trainiert, die öffentlich und spätestens am Tag des Earnings Calls verfügbar sind. Wir inkludieren Managementprognosen, da diese zu den öffentlichen Informationen zählen (siehe Kapitel 4 für Ergebnisse exklusive Managementprognosen). Die Decision-Tree-Verfahren entscheiden dabei automatisch, ob Managementprognosen ein relevantes Feature sind oder durch andere Features substituiert werden.

Optimierung der Modelle

Um die Prognosegenauigkeit zu erhöhen, führen wir vier Optimierungsschritte aus. Als erstes nutzen wir den Boruta-Algorithmus zur Feature Selektion (vgl. Kursa and Rudnicki 2010). Dadurch werden die ursprünglichen 91 Features auf die finalen 25 Features reduziert, da diese den Prognosefehler minimieren. Als zweites wenden wir sogenanntes „Feature Engineering" an. Dabei generieren wir aus den Prognosen des Managements zwei zusätzliche Informationen. Die erste Variable beinhaltet den prozentualen Prognosefehler der vorherigen Quartalsschätzung. Bei der zweiten Variable adjustieren wir die aktuelle Schätzung durch Multiplikation mit 1 minus dem letzten Schätzfehler in Betrag. Dadurch wird die aktuelle Managementschätzung weniger stark gewichtet, je höher der vorherige Prognosefehler des Managements ausgefallen ist. Außerdem benutzen wir drei kategorische Variablen (Firmen-Identifikationsnummer, Fama-French-Industriecode und den Quartalscode). Für diese kategorischen Variablen wenden wir „Target Encoding" an. Dabei sollen vergangene Werte der Zielvariable als Information in die kategorische Variable integriert werden.

Als drittes wenden wir „Model Blending" an. Dabei wird der finale Schätzwert aus einem gewichteten Durchschnitt von mehreren ML-Modellen ermittelt (im Folgenden „*lightGBM* Ensemble"), wodurch Verzerrungen eines einzelnen Modells reduziert werden. Die drei Modelle *lightGBM*, Extra Trees (eine leicht abgewandelte Form des Random Forest) und die Gradient-Boosted-Tree-Variante *xgboost* haben in unserem Datensatz aus 20 Modellen (bspw. MLP neuronales Netz und Lasso Regression) die höchste Genauigkeit. Die Gewichtung (die bei jedem Modell zwischen 0 und 1 liegt und in Summe 1 ergibt) bestimmen wir anhand der Kreuzvalidierung, die im nächsten Abschnitt erklärt wird. Als viertes ermitteln wir die optimalen Hyperparameter des *lightGBM*-Ensembles mit dem *Optuna*-Modul, das effizient verschiedene Hyperparameter-Kombinationen testet (vgl. Akiba et al. 2019).

Zeitreihen-Kreuzvalidierung

Standardmäßig wird beim ML der gesamte Datensatz in einen Trainings- und einen Testdatensatz aufgeteilt (vgl. Hastie et al. 2009). Um die Generalisierbarkeit der antrainierten Modelle auf bisher ungesehene Daten zu erhöhen (Minimierung des Overfitting), wird häufig der sogenannte „K-Fold Cross Validation"-Ansatz benutzt. Dabei wird der Datensatz in K (oftmals fünf) unterschiedlich angeordnete Blöcke geteilt, sodass K unterschiedliche Modelle antrainiert und K unterschiedliche Testdaten geschätzt werden. Da bei Umsatzprognosen ein Zeitreihenproblem vorliegt, verwenden wir eine vorwärts rollierende Kreuzvalidierung. Entsprechend wird sichergestellt, dass die Testdaten ausnahmslos zeitlich nach den Trainingsdaten liegen. Andernfalls würde es zu „Information Leakage“ kommen, da Zukunftsdaten zum Trainieren des Modells genutzt werden würden.

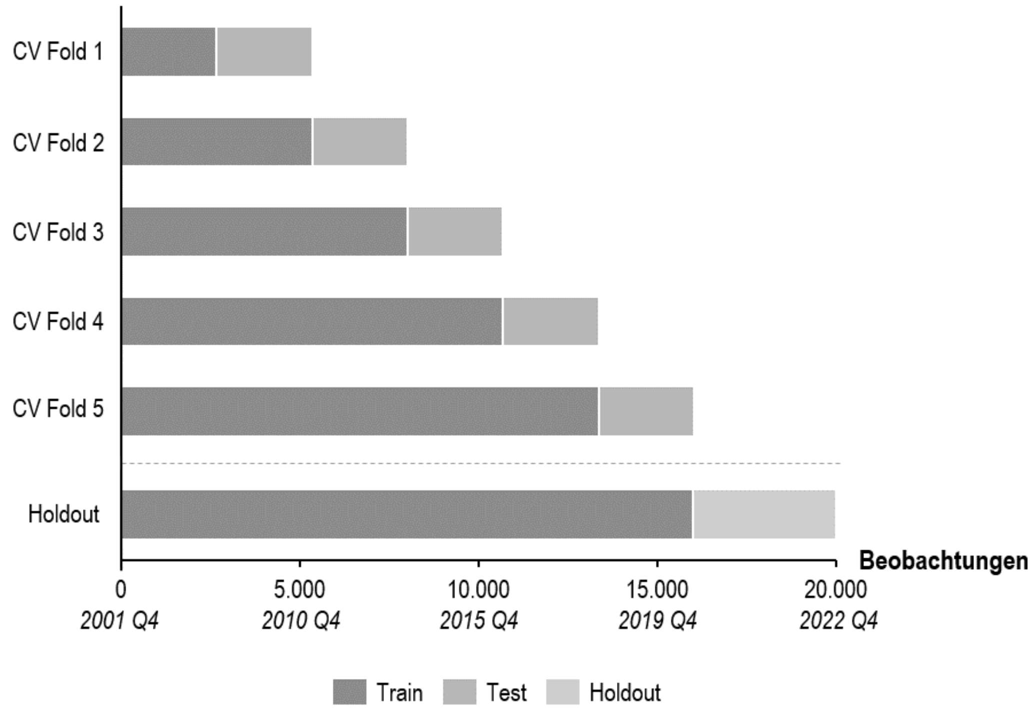


Abb. 2: 5-fache Zeitreihen-Kreuzvalidierung mit zusätzlichem Holdout-Set

Wir nutzen eine 5-fache Zeitreihen-Kreuzvalidierung mit zusätzlichem Holdout-Set (*Abb. 2*). Dadurch trainieren wir insgesamt sechs Modelle, die sechs unterschiedliche Umsatzzeiträume schätzen. Zunächst ordnen wir die 19.998 Quartalsbeobachtungen chronologisch an. Von diesen nutzen wir die ersten 80 % als Trainingsdaten und die letzten 20 % als einen Holdout-Datensatz, der unberührt von der Kreuzvalidierung ist. Das hat den Vorteil, dass wir die finalen Hyperparameter und die Ensemble-Gewichtung anhand der Kreuzvalidierung bestimmen und dann auf das unberührte Holdout-Set anwenden. Die Kreuzvalidierung erfolgt mit den vorher getrennten Trainingsdaten. Der erste Fold nutzt dabei die geringste Datenmenge mit jeweils 2.667 Beobachtungen in dessen Trainings- und Testdatensatz. Die genutzte Trainings-Datenmenge steigt mit jedem Fold, während die Anzahl an Testbeobachtungen konstant bei einem Wert von 2.667 bleibt. Bei der Kreuzvalidierung werden fünf überschneidungsfreie Zeiträume zwischen Q2/2007 und Q3/2020 prognostiziert. Beim Holdout-Set werden Umsätze im Zeitraum von Q3/2020 bis Q4/2022 geschätzt. Die optimalen Hyperparameter des *lightGBM*-Ensembles werden basierend auf dem minimalen durchschnittlichen Prognosefehler der fünf Folds bestimmt. Analog wird die optimale Gewichtung des Ensembles bestimmt, wovon *lightGBM* (*Extra Trees*; *xgboost*) im Optimum 58 % (37 %; 5 %) der Ensemble-Prognose ausmacht.

4. Ergebnisse der Machine-Learning-Forecasts und deren Praxisimplikationen

Abb. 3 zeigt die durchschnittlichen Prognosefehler (MAEs) des Managements- und des *lightGBM*-Ensemble-Modells für die sechs unterschiedlichen Zeiträume. Die Umsatzprognose des Managements liegt von Q2/2007 bis Q4/2022 im Durchschnitt 18,53 Mio. $ neben dem realisierten Quartalsumsatz. Die *lightGBM*-Ensemble-Modelle weisen hingegen einen niedrigeren Prognosefehler von 14,86 Mio. $ aus. Dadurch ergibt sich ein durchschnittlicher Prognosevorteil (vgl. *Ding* et al., 2020) von 19,6 % für das ML-Modell im Vergleich zu den Managementprognosen. Der Prognosefehler der ML-Modelle ist statistisch signifikant geringer als die Prognosefehler des Managements (p-Wert < 0,10 für alle Modelle). Auffallend ist, dass der erste Fold den geringsten Prognosevorteil vorweist. Das könnte dadurch erklärt werden, dass der Testzeitraum des ersten Folds die gesamte Finanzkrise abdeckt. Dennoch erreicht das ML-Modell auch in diesem Zeitraum einen statistisch signifikant niedrigeren Prognosefehler. Außerdem reduziert das Holdout-Modell den Prognosefehler während der Covid-19-Krise um 17,08 %. Es zeigt sich, dass die Prognosen auch in volatilen Krisenzeiten verbessert werden können.

Modell	Prognosezeitraum	Prognosefehler Management [Mio. $]	Prognosefehler lightGBM-ML-Ensemble [Mio. $]	Prognosevorteil Machine Learning
CV Fold 1	2007 Q2 - 2011 Q1	15,41	13,28	*+13,82%*
CV Fold 2	2011 Q1 - 2014 Q1	19,09	14,50	*+24,04%*
CV Fold 3	2014 Q1 - 2016 Q3	18,99	13,95	*+26,54%*
CV Fold 4	2016 Q3 - 2018 Q3	17,84	14,79	*+17,10%*
CV Fold 5	2018 Q3 - 2020 Q3	19,70	15,95	*+19,04%*
Holdout	2020 Q3 - 2022 Q4	20,14	16,70	*+17,08%*
Ø	**2007 Q2 - 2022 Q4**	**18,53**	**14,86**	***+19,60%***

Abb. 3: Prognosefehler des Managements- und des Machine-Learning-Modells der vierteljährlichen Umsätze

Aufgrund der Ausreißeranfälligkeit der Zielvariable bei Decision-Tree-Verfahren haben wir die 10 % umsatzstärksten Beobachtungen entfernt. Führt man die gleiche Analyse mit den 1.999 umsatzstärksten Beobachtungen durch, ergibt sich weiterhin ein Prognosevorteil (von 13,7 %).

Allerdings ist der Vorteil nicht mehr statistisch signifikant, was vermutlich an der geringen Stichprobengröße liegt. Die 1.999 Beobachtungen beziehen sich auf Beobachtungen des 90- bis 99-%- Perzentils (durchschnittlicher Quartalsumsatz von 4.801 Mio. $) und exkludieren das 99-%-Perzentil (durchschnittlicher Quartalsumsatz von 34.180 Mio. $), da diese Extremausreißer die Leaf-Nodes zu stark verzerren.

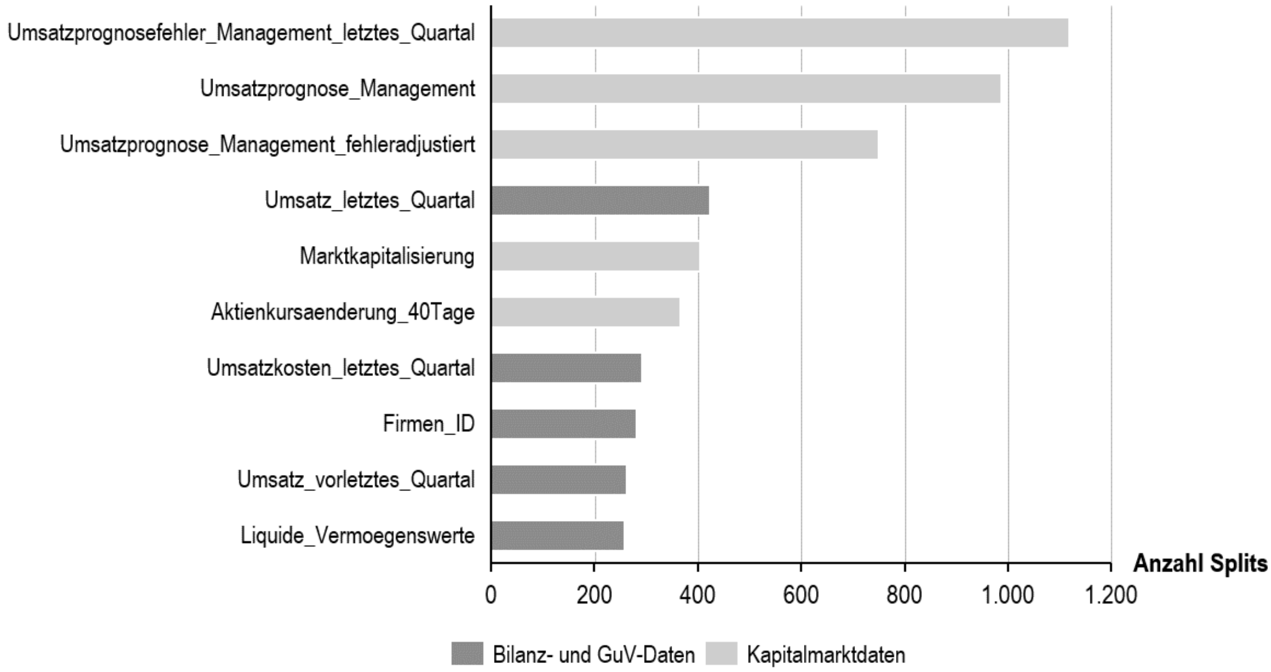


Abb. 4: Feature-Importance der genutzten Bilanz-, GuV- und Kapitalmarktdaten

Abb. 4 zeigt die **Relevanz der einzelnen Features** für das Holdout-Set („Feature-Importance"). Aus Übersichtlichkeitsgründen werden nur die zehn relevantesten Features abgebildet. Die Feature Importance wird in Abhängigkeit davon bestimmt, wie oft das jeweilige Feature als Split (zur Minimierung des Fehlerterms) in den Nodes der iterativen *lightGBM*-Bäume genutzt wird. Die drei wichtigsten Features beziehen sich auf die (fehleradjustierte) Umsatzprognose und den vorherigen Prognosefehler des Managements. Danach folgt der Umsatz des (vor)letzten Quartals sowie die Marktkapitalisierung und die Aktienkursänderung der letzten 40 Tage. Entfernt man die drei Managementprognose-Variablen, dann erzielen die Managementschätzungen einen Prognosevorteil von durchschnittlich 35,9 % gegenüber dem ML-Modell. Dieses Ergebnis wird im nächsten Abschnitt diskutiert.

Praxisimplikationen und Ergebnisdiskussion

Die Verbesserung der Prognosegenauigkeit durch Einsatz von ML basiert möglicherweise auf zwei Ursachen. Erstens werden nicht nur Zeitreihendaten eines Unternehmens genutzt, sondern ein Querschnitt aus ca. 1.000 Firmen. Obwohl die Zeitreihen einzelner Unternehmen nicht identisch sind, ähneln sich deren Muster teilweise (vgl. *Semenoglou* et al., 2021). Dadurch kann vor allem das Problem der geringen Datenverfügbarkeit behoben werden. Bei alleiniger Nutzung von Zeitreihendaten reduzieren sich die Beobachtungen eines Unternehmens in unserem Datensatz auf maximal 80 Beobachtungen. Dies würde die Generalisierbarkeit des antrainierten Modells stark reduzieren. Zweitens unterliegen klassische (menschliche) Forecasts Verhaltensverzerrungen. So werden Gewinnprognosen bei gutem Wetter unbewusst nach oben verzerrt (vgl. *Chen* et al., 2022) und Analysten übergewichten den vorherigen realisierten Wert in ihren Prognosen (vgl. Campbell and Sharpe 2009). Verzerrungen können allerdings auch bewusst vorgenommen werden. So neigen Manager teilweise dazu, ihre Forecasts zu optimistisch zu gestalten, damit sie ihre Aktienpakete profitabler veräußern können (vgl. Rogers and Stocken 2005). Durch Hinzufügen des vorherigen Prognosefehlers und der fehleradjustierten Prognose in die ML-Modelle werden die bewussten und unbewussten Verzerrungen in unseren Modellen merkbar reduziert.

Allerdings geht der Prognosevorteil verloren, wenn die (öffentlich verfügbaren) Managementprognosen aus dem Modell entfernt werden. Eine Erklärung hierfür ist, dass dem Modell relevante Input-Variablen fehlen. Entsprechend kann davon ausgegangen werden, dass das Management einen Informationsvorteil durch firmeninterne Daten hat, wie bspw. Lagerbestände, geplante Preisanpassungen und Investitionsentscheidungen. Durch Kombination der genannten ML-Vorteile (Informationszugewinn durch Querschnitts- und Zeitreihendaten und unverzerrte Informationsgewichtung) und der Nutzung firmeninternen Daten könnten Unternehmen potenzielle Steigerungen der Prognosegenauigkeit erzielen. Zum Beispiel konnte *Microsoft* durch Kombination von menschlichen Forecasts und ML-Verfahren den durchschnittlichen Prognosefehler der internen Forecasts von 2,7 % auf 1,5 % senken (vgl. *Wiprächtiger*, 2021). Sobald die relevanten Daten in strukturierter Form vorliegen und die ML-Modelle grundlegend erstellt sind, dauert ein Forecast in der Regel wenige Minuten bis wenige Stunden. Dadurch können potenzielle Zeit- und Kostenersparnisse realisiert werden.

Für die Ausführung von Predictive Analytics ist Expertenwissen im Unternehmen notwendig. So geht bei Nutzung der populären linearen Regression (inklusive Optimierung durch Lasso Regularisierung) der Prognosevorteil komplett verloren (54,3 % Prognosenachteil). Allerdings lässt sich der erzielte Prognosevorteil von 19,6 % des *lightGBM*-Ensembles nur durch Anwendung mehrerer Optimierungsschritte erzielen (siehe Abschnitt 3). Ohne Optimierungen reduziert sich der Prognosevorteil auf 1,7 % und verschwindet somit fast vollständig. Neben dem

Schritt der Modellerstellung und Ausführung sind noch weitere Prozesse notwendig, um das Forecasting skalierbar und automatisiert zu implementieren. Im Rahmen der „MLOps" müssen bspw. riesige Datenmengen zunächst möglichst effizient erhoben und abgespeichert werden.

Für Fachfremde wirken ML-Modelle und deren Ergebnisse oftmals als eine schwer nachvollziehbare „Black-Box" (vgl. Ding et al. 2020). Um diese Skepsis zu überwinden, ist eine schrittweise Implementierung in enger Kooperation zwischen Management, Controlling und Data Scientists (bzw. Machine Learning Engineers) ratsam. Zunächst sollten die menschlich generierten Forecasts durch ML-Prognosewerte unterstützt werden. Dazu sollte der Prognosefehler der menschlichen und ML-unterstützten Forecasts über mehrere Perioden verglichen werden und die Gewichtungen der beiden Komponenten adjustiert werden. Bei Zuverlässigkeit der ML-Prognosen kann zukünftig deren Gewichtung in den Umsatzprognosen gesteigert werden.

5. Fazit

Eine erhöhte Prognosegenauigkeit der Quartalsumsätze hat u.a. positive Auswirkungen auf den Planungs- und Budgetierungsprozess sowie auf Investitionsentscheidungen (vgl. Goretzki and Wiegmann 2022). Unser Beitrag beschäftigt sich mit der Frage, ob Umsatzprognosen des Managements durch den Einsatz von ML-Modellen verbessert werden können. Die Ergebnisse zeigen, dass die *lightGBM*-Ensemble-ML-Modelle den Prognosefehler über sechs verschiedene Zeitabschnitte von 2007 bis 2022 um durchschnittlich 19,6 % reduzieren können. Die Erhöhung der Prognosegenauigkeit erfolgt aufgrund des Informationszugewinns aus Zeitreihen- und Querschnittsdaten sowie der unverzerrten Informationsgewichtung der einzelnen Input-Variablen. Die Verbesserung der Prognosegenauigkeit gelingt dabei auch in volatilen Krisenzeiten. Neben dem Einsatz von ML zur Umsatzprognose existieren noch viele weitere sinnvolle Anwendungsgebiete in Unternehmen (bspw. Prognosen von Produktnachfrage, Kundenabwanderungen oder Lieferanten- und Kreditausfällen).

Zusammenfassend kann die Aussage getroffen werden, dass ML-Modelle die Genauigkeit von Managementprognosen signifikant erhöhen können. Entsprechend ist davon auszugehen, dass zukünftig immer mehr Unternehmen dem Beispiel von *Microsoft* (vgl. *Wiprächtiger*, 2021) folgen und den Prognoseprozess durch ML-Modelle unterstützen werden. Bei Vorhandensein von unternehmensinternen Daten (in strukturierter Form) kann zukünftig im Optimalfall das Forecasting zum Großteil von automatisierten ML-Prozessen übernommen werden. Ob der Prognoseprozess langfristig sogar komplett von künstlicher Intelligenz ausgeführt werden kann, bleibt eine offene Frage.

Literatur

- *Akiba, T./Sano, S./Yanase, T./Ohta, T./Koyama, M.,* Optuna: A Next-generation Hyperparameter Optimization Framework (2019).
- *Campbell, S. D./Sharpe, S. A.,* Anchoring Bias in Consensus Forecasts and Its Effect on Market Prices, in: Journal of Financial and Quantitative Analysis, 44. Jg. (2009), H. 2, S. 369–390.
- *Chen, C./Chen, Y./Pittman, J. A./Podolski, E. J./Veeraraghavan, M.,* Emotions and Managerial Judgment: Evidence from Sunshine Exposure, in: The Accounting Review, 97. Jg. (2022), H. 3, S. 179–203.
- *Chen, X./Ha Cho, Y./Dou, Y./Lev, B.,* Predicting Future Earnings Changes Using Machine Learning and Detailed Financial Data, in: Journal of Accounting Research, 60. Jg. (2022), H. 2, S. 467–515.
- *Ding, K./Lev, B./Peng, X./Sun, T./Vasarhelyi, M. A.,* Machine learning improves accounting estimates: evidence from insurance payments, in: Review of Accounting Studies, 25. Jg. (2020), H. 3, S. 1098–1134.
- *Geertsema, P. G./Lu, H.,* Relative Valuation with Machine Learning, in: Journal of Accounting Research, 61. Jg. (2023), H. 1, S. 329–376.
- *Goodman, T. H./Neamtiu, M./Shroff, N./White, H. D.*, Management Forecast Quality and Capital Investment Decisions, in: The Accounting Review, 89. Jg. (2014), H. 1, S. 331–365.
- *Goretzki, L./Wiegmann, L.,* Glaubwürdigkeit signalisieren: Die Rolle von Business Unit Controllern in Forecast Review Meetings, in: Controlling, 34. Jg. (2022), H. 4, S. 4–10.
- *Hastie, T./Tibshirani, R./Friedman, J.,* The elements of statistical learning: data mining, inference and prediction, 2. Aufl. 2009.
- *Hui, K. W./Matsunaga, S. R.,* Are CEOs and CFOs Rewarded for Disclosure Quality?, in: The Accounting Review, 90. Jg. (2015), H. 3, S. 1013–1047.
- *Ittner, C. D./Michels, J.,* Risk-based forecasting and planning and management earnings forecasts, in: Review of Accounting Studies, 22. Jg. (2017), S. 1005–1047.
- *Ke, G./Meng, Q./Finley, T./Wang, T./Chen, W./Ma, W./Ye, Q./Liu, T.-Y.,* LightGBM: A Highly Efficient Gradient Boosting Decision Tree, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA 2017, S. 3149–3157.

- *Kursa, M. B./Rudnicki, W. R.,* Feature Selection with the Boruta Package, in: Journal of Statistical Software, 36. Jg. (2010), H. 11, S. 1–13.
- *Makridakis, S./Spiliotis, E./Assimakopoulos, V.,* M5 accuracy competition: Results, findings, and conclusions, in: International Journal of Forecasting, 38. Jg. (2022), H. 4, S. 1346–1364.
- *Rogers, J. L./Stocken, P. C.,* Credibility of Management Forecasts, in: The Accounting Review, 80. Jg. (2005), H. 4, S. 1233–1260.
- *Semenoglou, A.-A./Spiliotis, E./Makridakis, S./Assimakopoulos, V.,* Investigating the accuracy of cross-learning time series forecasting methods, in: International Journal of Forecasting, 37. Jg. (2021), H. 3, S. 1072–1084.
- *Wiprächtiger, R.,* Digitale Transformation der Finanzfunktion am Beispiel von Microsoft, in: Controlling, 33. Jg. (2021), H. 2, S. 30–39.

Anmerkungen

Wir danken Sophie Kulig für ihre hilfreichen Verbesserungsvorschläge.

Stichwörter

Künstliche Intelligenz # Machine Learning # MLOps # Predictive Analytics # Umsatzprognosen

Keywords

artificial intelligence # machine learning # MLOps # predictive analytics # revenue forecasting

Summary

Accurate revenue forecasts are important for firms to minimize uncertainties in their planning and budgeting processes. The article shows that machine learning models increase revenue forecast accuracy by 20 % in comparison to management forecasts. In addition to increased forecasting accuracy, potential time and cost savings can be leveraged.

Implikationen für die Praxis

- Umsatzprognosen können durch den Einsatz von Machine-Learning-Modellen signifikant verbessert werden, wodurch Unsicherheiten in Planungs- und Budgetierungsprozessen reduziert werden können und Kapitalmarktteilnehmer eine bessere Informationsgrundlage zur Verfügung steht.
- Um die besten Prognoseergebnisse zu erzielen, ist aktuell eine Mischform aus menschlichen und Machine-Learning-generierten Forecasts empfehlenswert.
- Zur Implementierung von automatisierten und skalierbaren Machine-Learning-Forecasts sollte der Aufbau der dazugehörigen „MLOps“ schrittweise unter enger Zusammenarbeit von Management, Controlling, Data Scientists und Machine Learning Engineers erfolgen.
- Aufgrund der Effizienz von Machine-Learning-Modellen können neben der Erhöhung der Prognosegenauigkeit auch potenzielle Zeit- und Kosteneinsparungen erzielt werden.

Zentrale Aussagen

- Im Vergleich zu Umsatzprognosen des Managements können Machine-Learning-Modelle den Prognosefehler im Zeitraum von 2007 bis 2022 um durchschnittlich 20 % reduzieren. Dabei werden ausschließlich öffentlich verfügbare Daten benutzt.
- Die Erhöhung der Prognosegenauigkeit erfolgt aufgrund des Informationszugewinns aus Zeitreihen- und Querschnittsdaten, sowie der unverzerrten Informationsgewichtung der einzelnen Input-Variablen.
- Klassische statistische Modelle, wie bspw. die lineare Regression, sind modernen Machine-Learning-Ansätzen bzgl. Prognosegenauigkeit größtenteils stark unterlegen.

Affidavit

I hereby declare that I completed the papers submitted and listed hereafter independently and with only those forms of support mentioned in the relevant paper or in the following supplementary list. When working with the authors listed, I contributed no less than a proportionate share of the work. In the analyses that I have conducted and to which I refer in the papers, I have followed the principles of good academic practice, as stated in the Statute of Justus Liebig University Giessen for Ensuring Good Scientific Practice.

____

Hannes Gerstel

Giessen, 19th July 2023

Submitted papers:

I Gerstel, H., Kreilkamp, N., Schmidt, M., Wöhrmann, A., *Does the Devil's Advocate Approach Mitigate Escalation of Commitment?* Working Paper

II Sabel, R., Gerstel, H., Wöhrmann, A., *Celebrating Failure – The Effects of Failure Awards on Risk-Taking and Escalation of Commitment*. Working Paper

III Ewelt-Knauer, C., Gerstel, H., Khaled, M., Wöhrmann, A., *What Motivates Independent Directors? The Influence of Director Incentives on Director Decisions*. Working Paper

IV Bausch, A., Brede, M., Gerstel, H., Wöhrmann, A., *Do Opposites Attract? The Effect of Cultural Distance on Mergers and Acquisitions: Evidence from Glassdoor Reviews*. Working Paper

V Gerstel, H., Khaled, M., *Machine Learning vs. Management Forecasts: Können Machine-Learning-Modelle die Genauigkeit von Umsatzprognosen verbessern?* Controlling, 35. Jg. (2023), H. 4, https://doi.org/10.15358/0935-0381-2023-4