JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN

Dissertation

Essays on Management Accounting: A Behavioral Perspective on Management Control Systems and Firm Performance

Author Rebecca Sabel, MBA & M.Sc.

Submitted to Justus Liebig University Giessen
Department of Business Administration and Economics

Submission date May 27, 2025

Supervisors

Prof. Dr. Arnt Wöhrmann
Chair of Managerial Accounting,
Justus Liebig University Giessen

Prof. Dr. Corinna Ewelt-Knauer
Chair of Financial Accounting,
Justus Liebig University Giessen

Table of Contents

<table>
<tr><th>Study no.</th><th colspan="4">1</th></tr>
<tr><td>Title</td><td colspan="4">Celebrating Failure – The Effects of Failure Awards on Risk-Taking and Escalation of Commitment</td></tr>
<tr><td>Authors</td><td colspan="4">Rebecca Sabel, Hannes Gerstel, Arnt Wöhrmann</td></tr>
<tr><td rowspan="8">Author contribution</td><td></td><td>Sabel</td><td>Gerstel</td><td>Wöhrmann</td></tr>
<tr><td>Numeric share</td><td>0.6</td><td>0.3</td><td>0.1</td></tr>
<tr><td>Conceptual development of research question</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Development of theory</td><td>✓</td><td></td><td></td></tr>
<tr><td>Methodology</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>Acquisition of data</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>Analysis/interpretation of data</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>Writing the manuscript</td><td>✓</td><td></td><td>✓</td></tr>
<tr><td>Publication status</td><td colspan="4">Revise & Resubmit (R&R)
Contemporary Accounting Research (VHB-Rating 2024: A+)</td></tr>
<tr><td rowspan="4">Peer-reviewed conferences</td><td>2022</td><td colspan="3">Canadian Academic Accounting Association (CAAA) Annual Conference 2022 (Saskatoon)</td></tr>
<tr><td>2022</td><td colspan="3">American Accounting Association (AAA) Annual Meeting 2022 (San Diego)</td></tr>
<tr><td>2022</td><td colspan="3">European Network for Experimental Accounting Research (ENEAR) Conference and Doctoral Colloquium 2022 (Sevilla)</td></tr>
<tr><td>2023</td><td colspan="3">Experimental Research in Management Accounting (EXRIMA) Conference 2023 (Giessen)</td></tr>
<tr><td>Research approach</td><td colspan="4">Experimental study</td></tr>
<tr><td>Language</td><td colspan="4">English</td></tr>
<tr><td>Abstract</td><td colspan="4">Innovations and efficient resource allocations are essential for firm success. However, managers' "fear of failure" often prevents firms from achieving these goals. To counteract this fear, firms have started granting Failure Awards. Failure Awards reward managers who initiate a promising but risky idea or project that eventually had to be terminated when failure became imminent. In this study, we examine whether Failure Awards promote risk-taking and simultaneously reduce resource wastage by mitigating escalation of commitment (EoC). We conduct an experiment in which we manipulated whether a Failure Award was present or absent. In the Failure Award present treatments, we manipulated whether the Failure Award emphasized risk-taking (innovation-type Failure Award) or the early termination of failing projects (discontinuation-type Failure Award). In line with our predictions, we find that Failure Awards increase risk-taking, irrespective of the type. Furthermore, we find that EoC is significantly reduced if the Failure Award emphasizes discontinuation but not if it promotes risk-taking.</td></tr>
</table>

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

Rebecca Sabel
Giessen University (Germany)
Rebecca.Sabel@wirtschaft.uni-giessen.de

Hannes Gerstel
Giessen University (Germany)
Hannes.Gerstel@wirtschaft.uni-giessen.de

Arnt Wöhrmann
Giessen University (Germany)
Arnt.Woehrmann@wirtschaft.uni-giessen.de
Licher Str. 62,
35394 Giessen, Germany

Abstract

Innovations and efficient resource allocations are essential for firm success. However, managers' "fear of failure" often prevents firms from achieving these goals. To counteract this fear, firms have started granting Failure Awards. Failure Awards reward managers who initiate a promising but risky idea or project that eventually had to be terminated when failure became imminent. In this study, we examine whether Failure Awards promote risk-taking and simultaneously reduce resource wastage by mitigating escalation of commitment (EoC). We conduct an experiment in which we manipulated whether a Failure Award was present or absent. In the Failure Award present treatments, we manipulated whether the Failure Award emphasized risk-taking (innovation-type Failure Award) or the early termination of failing projects (discontinuation-type Failure Award). In line with our predictions, we find that Failure Awards increase risk-taking, irrespective of the type. Furthermore, we find that EoC is significantly reduced if the Failure Award emphasizes discontinuation but not if it promotes risk-taking.

Keywords: Escalation of commitment, Failure Award, fear of failure, psychological safety, risk-taking

JEL: M41

Data Availability: Contact the corresponding author.

We thank Jeremiah Bentley, Willie Choi, Victor van Pelt, Todd Thornock, and participants of the 2023 EXRIMA Conference, 2022 CAAA Annual Meeting, 2022 AAA Annual Meeting, and the 2022 ENEAR Conference and Doctoral Colloquium. We appreciate the financial support provided by The Chartered Institute of Management Accountants (CIMA).

I. Introduction

Failure is ubiquitous in organizations and often unavoidable on the path to success. However, employees often exhibit a fear of failure. Survey results indicate that 40% of employees are afraid of failure and thus spend 20-40% of their time worrying about making mistakes (Brassey et al., 2019). From a firm perspective, this fear of failure has a negative impact on firm performance.

First, the fear of failure prevents innovation despite its importance for firms' growth, efficiency, and productivity (Birkinshaw & Haas, 2016a). Innovations are subject to high uncertainty and closely linked to failure (Fischer et al., 2018). For instance, new product developments exhibit failure rates of 40% on average (Knudsen et al., 2023). Employees fear such high failure rates and thus exhibit risk-averse behavior to avoid the negative repercussions of failure, e.g., for their career or reputation (García-Granero et al., 2015; Wu, 2008; Zhou & George, 2001). In this vein, in a survey by the Boston Consulting Group, 31% of the respondents identify a risk-averse culture as a key obstacle to innovation (Birkinshaw & Haas, 2016b). Risk aversion gives rise to opportunity costs for (risk-neutral) shareholders if risk prevents managers from investing in projects with high expected returns (Baysinger et al., 1991; Eisenhardt, 1989; Wiseman & Gomez-Mejia, 1998).

Second, the fear of failure fosters escalation of commitment (EoC) (Johnson, 2017). EoC is a cognitive bias also known as "[o]ne of the most robust and costly decision errors" (Sleesman et al., 2012). It occurs when decision-makers 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). Employees afraid of failure hesitate to admit that it was a mistake to have started the (failing) project in the first place. Thus, to prevent image loss, they continue investing and hope for a return to profitability (Edmondson, 2003; Sleesman et al., 2012).

As the fear of failure may prevent innovation and increase escalation, management controls that attenuate such negative effects are required. In practice, an increasing number of firms have moved away from only rewarding success and have started to also 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 a merely symbolic financial reward and rely on "celebrating failure", e.g., by granting awards to employees during official ceremonies (Johnson, 2017; Supercell, 2021; TATA, 2021). Astro Teller, the director of Google's R&D division, "Google X", explains why Google uses Failure Awards as follows: "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 thus underlines the two goals of Failure Awards: (1) to encourage risk-taking and thus innovation and (2) to save resources via the early termination of failing projects (Johnson, 2017; Leber, 2016).

Whether Failure Awards can achieve these goals is an open question, as there is no empirical evidence of their effectiveness. This is where our study intends to contribute. We conduct an experiment to investigate whether Failure Awards can be used as a management control to promote risk-taking and reduce EoC.

Anecdotal evidence suggests that several types of Failure Awards exist. While the criteria for receiving a Failure Award (i.e., (i) risk-taking, (ii) failure, (iii) deliberate discontinuation) are usually identical, some Failure Awards put more emphasis on taking risk (*innovation-type* Failure Award), while others emphasize the timely termination of failing

[1] For instance, the marketing and communication agency Hill Holiday grants the "Epic Fail Award" (Proulx, 2019). Proctor & Gamble has 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 to 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 have been discontinued with beer and champagne (Deutschman, 2004).

projects (*discontinuation-type* Failure Award).[2] We therefore focus on the two endpoints of this continuum, i.e., *innovation-type* and *discontinuation-type* Failure Awards, and examine their effects on risk-taking and EoC.

We argue that both types of awards reduce the fear of failure by inducing psychological safety. Psychological safety is the feeling of safety that enables interpersonal risk-taking (Edmondson, 1999). This feeling makes decision-makers less afraid of the negative consequences of failure for their image or career. Thus, they are more willing to start risky projects. Consequently, hypothesis 1a (hypothesis 1b) predicts that employees take more risks when discontinuation-type (innovation-type) Failure Awards are present rather than absent. Notably, innovation-type Failure Awards explicitly encourage decision-makers to feel safe to experiment and take risks. Whether this more direct emphasis on experimentation and risk-raking results in more risk-taking under innovation-type instead of discontinuation-type Failure Awards leads to our first research question (RQ1).

As the fear of failure may also lead to escalation of commitment, we examine whether Failure Awards also reduce EoC if psychological safety is established. Both innovation-type and discontinuation-type Failure Awards make decision-makers feel safe to accept failure, as they do not anticipate the negative consequences of failure. Thus, they are more willing to discontinue a failing project. While this leads to a clear prediction for lower EoC in the case of Failure Awards highlighting discontinuation (H2), the effect of innovation-type Failure Awards is less clear, as a second effect must be considered. As mentioned above, the innovation-type Failure Award explicitly encourages employees to experiment. Thus, they may also hold on to a failing project, as they take the risk of betting on the small chance of turning the project

---

[2] For example, Hill Holliday has 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 grants 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 early termination of a failing project. The "Innovation Award" at 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).

profitable. Due to this opposing effect, we pose a research question on whether Failure Awards highlighting innovation effectively reduce EoC (RQ2).

To test our predictions and answer our research questions, we employ a 2×1+1 between-subjects experimental design.[3] 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 on two levels (*innovation-type* vs. *discontinuation-type* Failure Award). The operationalization is derived from practical examples of Failure Awards that either emphasize the importance of taking risks and innovating or of stopping resource wastage in failing projects.

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

In line with our predictions in H1a and H1b, we find that Failure Awards increase risk-taking irrespective of their type. However, we do not find that risk-taking is higher for innovation-type Failure Awards than for discontinuation-type Failure Awards, which answers our first research question (RQ1). Furthermore, we find that EoC decreases when discontinuation-type Failure Awards are used (H2). Regarding our second research question (RQ2), innovation-type Failure Awards do not reduce EoC. Accordingly, only discontinuation-type Failure Awards have a de-escalating effect.

[3] The research was conducted in an ethical manner. Specifically, subjects were treated anonymously in accordance with the relevant data protection regulations and were not exposed to specific risks. Furthermore, subjects were not deceived in any way or at any time. The institution at which the study was conducted does not have a review board to provide ethical clearance.

Additional analyses reveal the drivers of the distinct effects of innovation-type versus discontinuation-type Failure Awards on EoC. Notably, psychological safety is the key driver of Failure Award effectiveness. Psychological safety builds the feeling of being safe to admit mistakes, which reduces EoC for both types. However, if *innovation-type* Failure Awards are granted, the feeling of being safe to experiment is triggered, which offsets the EoC reducing effect. As this offsetting effect only materializes under innovation-type and not under discontinuation-type Failure Awards, only discontinuation-type Failure Awards reduce EoC.

To the best of our knowledge, this study is the first to examine Failure Awards empirically, thereby contributing to both management accounting practice and theory. First, from a practical perspective, we explore the idea of rewarding failure, which has become increasingly popular 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. Therefore, our study responds to the call of Cronin et al. (2021) for more research on the effect of communicating failure tolerance to employees.

Second, we contribute to the related literature on the effects of an open error management culture (EMC).[4] An EMC is the set of shared beliefs, norms, and common practices on how errors are addressed in an organization (van Dyck et al., 2005). Failure Awards can thus be regarded as a specific instrument of an open EMC, as failure is perceived as an acceptable outcome and an opportunity to learn from in both cases (Fischer et al., 2018; Gold et al., 2014; van Dyck et al., 2005). However, there is an important difference. Failure Awards not only communicate that failure is tolerated but also actively reward failure if a failing project is actively terminated by an employee. This difference might be important, as research on open EMC provides mixed findings on its effectiveness in reducing EoC. Whereas some papers find that a failure-tolerating culture decreases EoC (e.g., Simonson & Staw, 1992), others document

[4] Some studies also employ the term “error management climate”. As “culture” and “climate” are inherently difficult to differentiate, both concepts are treated interchangeably in this study.

an increase therein (e.g., Barton et al., 1989).[5] A potential explanation for these controversial findings is that project termination is not explicitly incentivized in an open EMC. Failure Awards, instead, not only signal failure tolerance but also reward failure if failing projects are terminated by employees. Thus, Failure Awards may effectively decrease EoC.

Third, our findings have important implications for the design of (non-monetary) incentive schemes using Failure Awards. Specifically, we show that for promoting innovations and risk-taking, the type of Failure Award is irrelevant, while it matters for EoC. Our results show that only Failure Awards emphasizing project termination significantly reduce EoC. This is important, as many firms today are using innovation-type Failure Awards and thereby neglect the possible benefits of simultaneously reducing EoC. Some examples are the "Epic Fail Award" by Hill Holiday (Proulx, 2019), the "Heroic Failure Award" by Proctor & Gamble (Anthony, 2020), and the "Lean Forward; Fail Smart Award" (NASA, 2021).

From a theoretical perspective, we also contribute to research on intangible rewards and social recognition. While there is a vast literature on tangible rewards (e.g., Cardinaels et al., 2021; Choi & Presslee, 2023; Heninger et al., 2019; Jeffrey, 2009; Kelly et al., 2017), intangible rewards have received much less attention. Studies on recognition programs providing symbolic rewards find positive effects on performance and effort (e.g., Lourenço, 2016; Wang, 2017) and, under certain circumstances, on creative performance (Huo, 2020). However, these studies focus on the recognition of success (not failure) and usually adopt social-comparison theory to predict that employees strive for a positive self-image; thus, even non-monetary incentives are deemed effective (Tafkov, 2013). Similar research on failure is lacking. Hence, we add a new perspective by not restricting recognition to "best performance" and successful outcomes but instead include the reward of failure.

[5] Barton et al. (1989) find that an open EMC increases participants' investments in a failing project. In their experiment, an open EMC was implemented by informing participants that their initial investment decision had demonstrated good judgment even though their project later threatened to fail. One potential explanation for their finding is that participants were more likely to accept failure through a decreased fear of failure and thus held on to failing projects. Failure Awards, instead, explicitly require project termination.

Moreover, we extend the research on risk-taking. We show that Failure Awards can overcome the prevailing risk aversion of decision-makers. Furthermore, we contribute to the accounting phenomenon in EoC research (Cheng et al., 2003; Mahlendorf, 2015). That is, we identify discontinuation-type Failure Awards as a new and cost-efficient debiasing tool that reduces EoC. Failure Awards require a rather low input of resources compared to other de-escalation strategies, e.g., hiring a third-party expert (Behrens & Ernst, 2014). Finally, we shed light on the psychological mechanism that reduces the fear of failure through psychological safety (Frazier et al., 2017). We show that the feeling of being safe to admit failure encourages risk-taking and reduces EoC. However, we also show that the effect of feeling safe to experiment, which is exclusively triggered by the innovation-type Failure Award, offsets this EoC-reducing effect of psychological safety.

II. Background and Hypothesis Development

Failure Awards and Psychological Safety

In this section, we define several constructs that are important for our theory. First, it is important to understand failure and its implications for employees. We define failure as a negative performance outcome due to a lack of success, “bad luck” or the inability to achieve a desired goal that causes employees to feel at least partially responsible for it (Cronin et al., 2021). Feeling responsible for failure results in the “fear of failure”, i.e., the “[…] disposition to avoid failure and/or the capacity for experiencing shame and humiliation as a consequence of failure” (Atkinson, 1957). This fear materializes because organizations typically strive for high performance by installing management processes based on predictability and efficiency, leaving little to no room for failure (Birkinshaw & 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.

To counteract the fear of failure, firms have started granting Failure Awards (Johnson, 2017; Kuvalekar & Ravi, 2019), rewarding them to employees who have shown their willingness to innovate and take risks but have failed. Failure Awards are often granted during award ceremonies that express a company's appreciation and are associated with no or a merely symbolic financial reward.

Failure Awards are thus intended to fulfill two goals at once: (1) encouraging innovation by making it safe to take risks and (2) saving resources by making it safe to admit failure 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 "L*ean 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, NASA uses an *innovation-type* Failure Award that emphasizes risk-taking. Notably, to receive this award, properly handling the failure (e.g., 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 the failing project (*discontinuation-type* Failure Award*)*.

Such practical examples show that firms use specific criteria to award Failure Awards (e.g., Google X (Johnson, 2017)). Based on these examples, we derive the following general 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 terminated the failing project in a timely manner. While these three criteria must be met for innovation- *and* discontinuation-type Failure Awards, the examples cited above show that some firms put more emphasis on the innovation criterion while others stress the discontinuation requirement. In any case, their employees do not qualify for a Failure Award without a deliberate and timely termination in case of project failure.

Moreover, Failure Awards induce psychological safety (Baer & Frese, 2003; Cannon &

Edmondson, 2005; Edmondson & Lei, 2014; James et al., 1977). In a psychologically safe environment, individuals feel safe to take interpersonal risks, as they do not fear any negative consequences for their status or career (Edmondson, 1999; Kahn, 1990). According to Edmondson (2003), this is important, as individuals make decisions by assessing the interpersonal risk (e.g., the risk of being perceived as incompetent) associated with a specific action. Failure Awards are thus a credible signal that failure does not result in adverse consequences to one's reputation or career. Failure Awards acknowledge the courage needed to engage in promising but risky endeavors and demonstrate the firm's appreciation. Consequently, employees feel psychologically safe, and their fear of failure is reduced. Hence, psychological safety encourages employees to admit failure. Through this feeling, individuals do not perceive their reputation or career to be at risk if a project fails.

The Effect of Failure Awards on Risk-Taking

In this section, we build on psychological safety to predict that Failure Awards increase risk-taking if either discontinuation-type (H1a) or innovation-type (H1b) Failure Awards are used.[6] Furthermore, we posit a research question on whether the risk-inducing effect of Failure Awards differs between these two types (RQ1).

Motivating risk-taking is important as agency theory assumes that agents (i.e., employees) are risk-averse (Eisenhardt, 1989; Wiseman & Gomez-Mejia, 1998). Individuals are therefore less likely to engage in risk-taking if they perceive their wealth to be at risk (Keil et al., 2000; Sitkin & Weingart, 1995; Wong, 2005). This feeling materializes when employees expect project failure. On the one hand, employees fear direct monetary consequences if their compensation is performance-contingent and a project fails. On the other hand, indirect (monetary and non-monetary) consequences may also occur. While Failure Awards do not

[6] While risk-taking can be easily described as the choice of a risky decision (Barki et al., 1993), defining risk is more complex. However, the various definitions of risk exhibit two similarities: (1) the probability that an undesirable outcome will occur and (2) the consequences thereof (e.g., losses or decreased returns) (Barki et al., 1993; Highhouse & Yüce, 1996; Sitkin & Pablo, 1992). Thus, risk is expressed through the variance of expected decision outcomes.

compensate for the direct monetary consequences of failure, they reduce the indirect negative consequences. Monetary indirect consequences arise if future career and promotion prospects are harmed in case of failure, while non-monetary consequences result from expected reputation and image loss (Hirshleifer, 1993). Hence, 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. Psychological safety lowers the level of perceived risk (Palanski & Vogelgesang, 2011) and, thereby, the concerns about the indirect consequences of failure.

To achieve this goal, Failure Awards focus on the end or outcome of an investment (or similar) project and explicitly communicate that any outcome, including failure, is acceptable. By granting a Failure Award, a firm signals that if failure materializes due to high risk, it is acceptable, and thus, individuals feel *safe to admit failure*. Accordingly, the fear of taking risks at the beginning of a project is reduced, and risk-taking is indirectly encouraged. We refer to this mechanism as the psychological safety factor (PS factor). This mechanism is the same for both Failure Award types, as the construct of the Failure Award, i.e., awarding a failing outcome, decreases the severity of failure in both cases. Hence, we predict that both types of Failure Awards increase risk-taking, which leads to hypotheses H1a and H1b:

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.

Instead of focusing on the end of an investment project and communicating that any outcome, including failure, is acceptable, one might focus on the beginning of the project. A firm may want to foster risk-taking by explicitly creating the feeling that employees are *safe to take risks and experiment*. We refer to this mechanism as the safety to experiment factor. The safety to experiment factor is exclusively triggered by the innovation-type Failure Award because this type directly encourages individuals to experiment and innovate. The question thus arises whether this additional effect leads to more risk-taking under innovation-type Failure Awards. This is only the case if the effect of the PS factor and the safety to experiment factor are additive. One reason why both effects might not be additive is that the utmost a firm can do to increase risk-taking is to communicate that all consequences of taking risks, including failure, are acceptable. This might bear more weight than simply emphasizing risk-taking. Consequently, we posit the following research question:

> *RQ1: Is risk-taking higher when innovation-type Failure Awards are granted than when discontinuation-type Failure Awards are granted?*

The Effect of Failure Awards on Escalation of Commitment

After selecting and initiating a project, managers often remain engaged even if the project is failing. Hence, we next discuss how Failure Awards affect EoC. While we derive a directional prediction for discontinuation-type Failure Awards, we pose a research question for innovation-based Failure Awards.

Staw and Ross (1987) identify project, psychological, social and structural drivers of EoC. As we explain below, Failure Awards affect EoC through psychological and social drivers. The psychological determinants can be explained using self-justification theory (Festinger, 1957; Sleesman et al., 2012). According to this theory, decision-makers feel the need to justify their initial decision to start a project if it performs poorly (Brockner, 1992; Sleesman et al., 2012). The sunk cost fallacy may even facilitate self-justification pressures, as decision-makers do not want to be perceived as resource wasters (Arkes & Blumer, 1985).[7] Hence, they escalate their commitment to avoid psychological costs in case of (project) failure.

The social determinants of EoC imply that others, such as evaluators or rivals, indirectly affect decision-makers (Sleesman et al., 2012; Staw & Ross, 1989). According to self-

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

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 their initial course of action (Edmondson, 2003; Sleesman et al., 2012). Thus, decision-makers stay committed to their initial decision.

Failure Awards induce psychological safety by signaling that project failure does not indicate poor performance of the decision-maker. Hence, Failure Awards reduce the pressure to justify why a failing project was initiated (self-justification pressure) and mitigate concerns of being perceived as incompetent (impression management concerns). Accordingly, employees do not fear any negative consequences for their image or career if they admit failure and terminate a project.

This line of thought is supported by Simonson and Staw (1992), who find that self-justification pressure can be decreased by informing participants that their previous decisions resulting in negative outcomes are not an indicator of their intelligence. Similarly, Heng et al. (2003) show that assuring decision-makers that their superior's opinion about them will not be affected by their project's outcome reduces EoC. Finally, Mahlendorf (2015) demonstrates that organizational allowance for failure reduces managers' perceived threat of project failure, which reduces EoC. Thus, psychological safety (PS factor) reduces decision-makers’ reluctance to terminate a failing project, and EoC is reduced.

While this allows a clear prediction of an EoC-reducing effect of discontinuation-type Failure Awards, the effect of innovation-type Failure Awards remains less clear. As discussed for RQ1, innovation-type Failure Awards trigger not only the PS factor but also the safety to experiment factor. While both effects, the PS factor and safety to experiment factor, work in the same direction for risk-taking, they work in opposite directions for EoC. Decision-makers who feel safe to experiment might be encouraged to stay committed to a failing project, as they do not expect any negative consequences in case of project failure. Thus, they might bet on the small chance to turn the failing project profitable by project continuation. For innovation-type

Failure Awards, it is therefore questionable whether the de-escalating effect of psychological safety (PS factor) is offset by the risk-encouraging effect of feeling safe to experiment (safety to experiment factor). Consequently, we posit a directional hypothesis for discontinuation-type Failure Awards and a research question for innovation-type Failure Awards:

H2: Escalation of commitment is lower when discontinuation-type Failure Awards are granted than when Failure Awards are not granted.

RQ2: Do innovation-type Failure Awards reduce escalation of commitment?

III. Research Design

Experimental Design and Procedure

To test our predictions and answer our research questions, we employ a 2×1+1 between-subjects experimental design. We manipulate the type of Failure Award on two levels, *innovation-type* and *discontinuation-type* Failure Award. Furthermore, a Failure Award absent treatment (control group) is employed in which Failure Awards are not provided. The experiment was programmed using oTree (Chen et al., 2016) and conducted online on Amazon Mechanical Turk (MTurk) due to the COVID-19 pandemic and social distancing restrictions.

Figure 1 depicts the experimental procedure. Before working on the main tasks where we measured risk-taking (risk task) and escalation of commitment (EoC task), participants had to pass an eligibility check designed for MTurk workers and participated in a lottery task to measure ex-ante risk preferences. At the end of the experiment, participants learned their compensation and responded to a post-experimental questionnaire (PEQ).

During the eligibility check, participants had to demonstrate their knowledge of the expected value calculation that was required for the main task. Only participants who successfully calculated the expected value of a prize wheel could proceed. Next, participants completed the lottery task to measure their ex-ante risk preferences. Similar to Sprinkle et al. (2008), 15 scenarios were presented. Each scenario consisted 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 indicated 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. A random mechanism in the experiment chose one of the 15 scenarios and determined whether participants received an additional compensation of $1.50, $0.75 or $0, depending on the participant's lottery choice.

[Insert Figure 1 about here]

Participants then learned that they would act as project managers at the fictitious company "CleverClean", and the main tasks were described. During the main task, participants needed to decide in which of two projects to invest (risk task) and whether to continue investing when failure became imminent (EoC task). Participants knew that in addition to the payment from the lottery task and a fixed payment ($1.00), they could earn a performance-contingent payment that depended on their decisions during the main task. Compensation and other financial information was provided in lira (the experimental currency).[8] The performance-contingent compensation was 1% of the project account balance, managed by the participants. At the beginning of the experiment, the project account was credited with 5 m lira. Investments reduced this amount, and proceeds from investments increased it.

Next, the manipulation (presence and type of Failure Award) was described, and participants had to pass a quiz to verify their understanding of the task, the compensation, and the manipulation.

Failure Award Manipulation

Participants in the Failure Award treatments learned that CleverClean's management had introduced Failure Awards, read an example of a recent award winner, and watched a video showing an excerpt from the award ceremony. In practice, there is no single name for "Failure

[8] At the end of the experiment, all lira earned by participants were converted into dollars at a rate of 20,000 lira per dollar.

Awards", but firms use various (unique) names (e.g., Epic Fail Award, Dare to Try Award, Heroic Failure Award, etc.). While we employ the term Failure Award in this paper, the experimental materials referred to the award as "Courage Award", which gives us the opportunity to define "courage" differently depending on the two Failure Award treatments (Figure 2).[9]

Participants in the innovation-type treatment were 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 project". In the discontinuation-type treatment, participants were 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).[10] Importantly, the criteria to receive a Failure Award were kept constant across the two Failure Award conditions. Participants receive a Failure Award only if they a) start a risky project and b) deliberately terminate the project as soon as c) failure becomes imminent.

[Insert Figure 2 about here]

Practical examples show that Failure Awards have a symbolic meaning and are often non-monetary (e.g., trophies, applause, award ceremonies) or have only a symbolic cash component (e.g., Google X) (Johnson, 2017; Stewart, 2015). Thus, participants in the Failure

9 A pretest revealed that the name "Failure Award" was perceived by some as undesirable due to the negative connotation of "failure". This concern typically does not arise in practice. For example, as the CIO of Hill Holliday who grants the Epic Fail Award states, "[d]espite its awful-sounding name, this award has become something that Hill Holliday employees strive to win." (Proulx, 2019). In an experimental setting, the possibilities of convincingly presenting the Failure Award to mitigate these concerns are—compared to a real firm setting—very limited. Accordingly, the experimental materials use the term "Courage Award".

10 To differentiate the provided Failure Award type manipulation from a goal-setting manipulation (Kachelmeier et al., 2016), all treatments receive information indicating 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 the appropriate behavior by additionally rewarding this behavior (Kachelmeier et al., 2016).

Award conditions learned that in addition to an award ceremony, award winners receive the symbolic amount of 2,000 lira, which equals $0.10 (approx. 2% of the average total compensation).

Risk Task

The risk task is similar to the choice problems from Kahneman and Tversky (1979). Participants had to choose between investing in project A (i.e., Smart Vacuum Robot) or B (i.e., Smart Mop Robot). While both projects were risky, project B was associated with higher risk, i.e., greater variance in expected cash flows, and a higher expected value compared with project A. The experimental materials informed participants that the company preferred projects with higher expected returns.[11] Thus, the riskier project, i.e., project B, was economically preferred compared to the safer project A.

Participants were provided with a brief description of the two projects, cash flow forecasts for a best-case and a worst-case scenario, and investment ratings. Figure 3 shows the experimental materials for project A, i.e., the Smart Vacuum Robot, and Figure 4 shows those for project B, i.e., the Smart Mop Robot. While the capital requirement for both projects was identical (3 m lira), the probabilities of the two scenarios and the expected cash inflows differed between the two projects. Participants were made aware that all financials were predicted values reflecting only the information available at that stage of the experiment. Participants could easily calculate that project B, i.e., the economically preferred project, was expected to yield a return (expected cash inflows less investment) of 4 m lira and a variance of 49 m. Project A, instead, promised a lower expected return of 3.5 m lira and a lower variance of 2.25 m. The obvious difference in variance allows a strong test of our theory.

Participant compensation was 1% of the final "project's account balance". This account

[11] This is also in line with expected utility theory, which suggests that (risk-neutral) individuals should make decisions based on expected returns and therefore always choose the option with higher expected returns independent of the inherent risk (i.e., variance) (Kahneman & Tversky, 1979; Schoemaker, 1982).

was credited with 5 m lira at the start of the experiment. During the experiment, cash outflows decreased and cash inflows increased the account balance. Selecting the economically preferred project B yielded an expected balance of 9 m lira (i.e., initial project account balance (5 m) + expected project return (4 m)). Project A promised 8.5 m lira (i.e., initial project account balance (5 m) + expected project return (3.5 m)). To summarize the financial information, a "star" rating was provided that visualized that project B implied more risk compared to A but offered higher expected returns. As firms in the real world provide Failure Awards not for every kind of failure but only if a project entailed a substantial amount of risk, participants were informed that only the riskier project (B) qualified for the Failure Award. We measure risk-taking—our first dependent variable—based on the project participants decided to invest in. Being involved in the investment decision increases personal responsibility, facilitating EoC (Denison, 2009; Schoorman & Holahan, 1996), which we discuss below.

[Insert Figures 3 and 4 about here]

Escalation of Commitment Task

After making the investment decision, participants were informed that 12 months had passed and that they would now receive an update about the selected project.[12] To induce an EoC setting, decision-makers must receive negative feedback on their initial decision and decide whether to keep investing in the failing project (Wong et al., 2006). For example, Seybert (2010), Brink et al. (2020), and Denison (2009) inform participants about a decline in expected cash flows of an investment project. Similarly, we informed participants that the project development was below expectations (e.g., lower expected sales due to a new competitor) and that an additional investment of 1 m lira was required to continue the project. The updated

[12] To measure the effect of Failure Awards on EoC, we are only interested in participants selecting the riskier project. This is because taking on a substantial amount of risk is a prerequisite to receiving a Failure Award. However, due to potential fairness concerns of MTurk workers, we also let the participants who invested in the safer project continue the experiment and paid them accordingly. For simplicity, we only describe the financial scenarios faced by the participants who invested in the riskier project during the EoC task in this section.

predicted financials indicated that the expected return if project B is continued would be 0.32 m lira instead of the initially expected 4 m lira. As participants' performance-based compensation is 1% of the final project account, they could expect to earn 53,200 lira (i.e., 1% × [initial project account balance (5 m lira) + expected return (0.32 m lira)]) when continuing project B, compared to initially 90,000 (i.e., 1% × [initial project account balance (5 m lira) + expected return (4 m lira)]).

Alternatively, participants could terminate the project and invest the 1 m lira in an alternative project that promised an even higher expected return (Brink et al., 2020; Seybert, 2010). If participants decided on this option, they could expect a performance-contingent compensation of 60,000 lira, which exceeds the expected compensation of 53,200 lira when investing in the failing project. Thus, terminating the failing project and investing in the alternative investment project was the economically preferred option.

Participants in the two Failure Award conditions were reminded that they would (definitely) receive a Failure Award of 2,000 lira if they decided to invest in the alternative project. The 2,000 lira award is included in the amount of 60,000 lira from above, i.e., the expected compensation when investing in the alternative project (i.e., 1% × [initial project account balance (5 m lira) + expected project value (0.8 m lira)] + Failure Award (2,000 lira)).

Participants in the Failure Award absent condition naturally received no Failure Award if they decided to terminate the failing project and invest in the alternative project. To hold all treatments economically equivalent and to rule out that the compensation associated with the Failure Award drives our effects, the expected cash inflow of the alternative project was higher (5 m instead of 4.8 m lira) to compensate for the lack of Failure Award.[13] This resulted in an identical expected compensation in the Failure Award absent condition (i.e., 1% × [initial

[13] To rule out the role of the symbolic compensation associated with the Failure Award in our effects, we also included the following item in the post-experimental questionnaire: "The monetary compensation of $0.10 (2,000 lira) from the Failure Award was important to me." On average, participants responded 2.8 on a Likert scale ranging from 1 (not important) to 7 (highly important). As 2.8 is significantly below the scale's midpoint ($p < 0.01$), we thus conclude that it is not the compensation that drives our effects.

project account balance (5 m lira) + expected project value (1 m lira)]). Table 1 summarizes participants' expected compensation dependent on their continuation decision and the treatment.

[Insert Table 1 about here]

To measure EoC, i.e., our second dependent variable, participants had to recommend to the management on a scale from 0 to 100 whether to continue the failing project or to invest in the alternative investment opportunity (0% = definitely terminate, 100% = definitely continue) (Keil et al., 2000; Wong, 2005). To create impression management concerns and self-justification pressure, participants were told that their decisions would be reviewed and that they might receive written feedback. Approximately 5% of the participants were randomly selected and received a message through MTurk with feedback on the rationality of their decisions.

While such a recommendation decision allows a fine measurement of EoC, a binary decision is required to decide whether the termination criterion of the Failure Award is met. Thus, participants were asked whether to continue or terminate the failing project. Participants knew that if they decided to continue, they would need to make another decision in 12 months. If participants decided to terminate their project, the task ended. Otherwise, they entered a second round. We included this second round, as Brockner (1992, p. 40) argues that "[…] escalation situations include repeated (rather than one-shot) decision-making in the face of negative feedback […]". Participants could delay the termination decision in the first EoC round and justify this by relying on the (small) chance to turn the project profitable and if not to end it if it continues to fail. The second EoC round was almost identical to the first round, but outcome probabilities worsened again (Behrens & Ernst, 2014) and the Failure Award was granted only with a 50% probability.[14]

[14] This reduced likelihood is implemented because a delayed project termination contradicts the objective of a Failure Award.

For our tests, we rely only on the first (fine) EoC measurement, as project termination is the economically preferred decision at this point. Our main results are inferentially identical if we use the binary EoC measure. In the additional analyses, we discuss participants who decided to terminate the project in the second instead of the first round.

Participants

Participants were recruited from Amazon MTurk through a publicly advertised human intelligence task (HIT). The primary reason for using MTurk is that the COVID-19 pandemic did not allow a laboratory experiment. 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 & 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 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. Furthermore, using mobile devices for the task was prohibited to minimize possible distractions. Last, if participants spent less than the bare minimum of required time on a page based on minimal page times collected during the pretest, they could not proceed with the experiment (Hunt & Scheetz, 2019).

In total, 277 persons participated in the study. Of these, 13 participants had to be excluded because they failed at least one attention check. The remaining sample is therefore 264 participants. The participants' average age was 40.3 years, 37.12% 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 63.64% of the participants were risk-averse, 16.67% were risk-neutral, and 19.70% were risk-seeking. This is in line with previous research that finds a preference for risk aversion among individuals (Crosetto & 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 Failure Awards (all p-values > 0.21).[15] Hence, randomization was successful.

At the end of the experiment, participants learned their compensation. For participants who continued the project, a random mechanism determined whether the best-case scenario or the worst-case scenario of the project materialized. 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).

IV. Results

Comprehension and Other Checks

Before testing our hypotheses, we verified the participants' correct understanding of the task. To create a valid EoC setting, participants needed to comprehend that their initially chosen project was failing. Thus, we asked 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

[15] All p-values are reported as two-tailed unless stated otherwise.

disagree, 7 = totally agree). The mean value is significantly above the scale midpoint ($p < 0.01$). Hence, subjects understood that their project was failing. Furthermore, we check whether Failure Awards created a culture in which participants perceived that failure was tolerated (i.e., open EMC). On a 7-point Likert scale, we find that subjects in the Failure Award treatments agreed 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 created an open EMC. Moreover, we check whether participants correctly identified the risker project. Participants indicated which project they believe to be riskier on a 7-point scale (1 = Smart Vacuum Robot, 7 = Smart Mop Robot). With a mean value of at least 6.36, participants correctly identified the Smart Mop Robot as the riskier project in all treatments.[16] Last, participants in both Failure Award conditions correctly selected on their first try the three conditions required to qualify for a Failure Award (i.e., 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.

Descriptive Results and Hypotheses Tests

Table 2, Panel A and Figure 5, Panel A illustrate the descriptive statistics for risk-taking. Risk-taking means the percentage of participants in a treatment that decided to invest in the riskier project. H1a predicts that risk-taking is higher when discontinuation-type Failure Awards are granted than when Failure Awards are not granted. Consistent with H1a, more participants invested in the riskier project when discontinuation-type Failure Awards were present (74%) rather than absent (47%).

[Insert Table 2 and Figure 5 about here]

An analysis of variance (ANOVA) shows that risk-taking significantly differs among the three treatments (Table 3, Panel A, $F = 10.18$, $p < 0.01$). For the formal test of H1a, we apply

[16] Excluding the 11 subjects who indicated a value of 4 or less leads to inferentially identical results.

pairwise comparisons. The results in Table 3, Panel B show that participants in the discontinuation-type treatment are more likely to invest in the riskier project than participants in the Failure Award absent treatment ($t = 3.83$, $p < 0.01$).[17] Hence, H1a is supported.

H1b predicts that risk-taking is higher when innovation-type Failure Awards are granted than when Failure Awards are not granted. In line with H1b, the descriptive results show that risk-taking is higher when innovation-type Failure Awards are present (73%) than when Failure Awards are absent (47%). The pairwise comparisons in Table 3, Panel B show that risk-taking is significantly higher in the innovation-type Failure Award condition ($t = 3.78$, $p < 0.01$). Thus, H1b is supported.

The first research question (RQ1) investigates whether risk-taking is higher for innovation-type Failure Awards than for discontinuation-type Failure Awards. The descriptive results show that risk-taking under innovation-type Failure Awards (73%) and discontinuation-type Failure Awards (74%) is almost identical. The difference is not statistically significant ($t = 0.11$, $p = 0.91$). We refer to this finding in the additional analyses section.

[Insert Table 3 about here]

Next, we focus on EoC. Only participants who selected the riskier project could receive a Failure Award. Thus, we use only these participants when we examine the effect of Failure Awards on EoC. Thus, the sample size decreased from 264 to 165 participants. Figure 5, Panel B illustrates the results.

H2 predicts that EoC is lower when discontinuation-type Failure Awards are granted than when Failure Awards are not granted. The descriptive results in Table 2, Panel B show that the likelihood of project continuation is lower when discontinuation-type Failure Awards are

[17] To estimate treatment effects on binary outcomes, applying linear OLS regression models is generally more appropriate than using logit models (Gomila, 2021). Taking the discontinuation-type as the baseline in a linear regression model (untabulated), risk-taking in the discontinuation treatment significantly differs from the Failure Award absent treatment ($t = 3.83$, $p < 0.01$) but does not differ from the innovation-type treatment ($t = -0.11$, $p = 0.910$). Alternative logit regressions confirm these results ($z = 3.53$, $p < 0.01$; $z = -0.12$, $p = 0.905$).

granted (43.96%) than when Failure Awards are absent (62.68%). This is consistent with H2. The ANOVA results in Table 4, Panel A show that EoC significantly differs across all three treatment groups ($F = 3.16$, $p = 0.045$). To formally test H2, we use pairwise comparisons (Table 4, Panel B). Our results show that EoC is significantly lower in the discontinuation-type Failure Award treatment 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 difference is not statistically significant (Table 4, Panel B, $t = -1.28$, $p = 0.201$). We discuss this finding in the additional analyses section.

[Insert Table 4 about here]

Additional Analyses

This subsection provides further analyses and explores questions from the post-experimental questionnaire to further test the theory that underlies our hypotheses. First, we focus on psychological safety (PS factor) and the safety to experiment factor, which are core to our theory. Next, we examine whether Failure Awards mitigate the sunk cost fallacy, i.e., an important driver of EoC. Finally, we examine EoC behavior in the second EoC round.

Factor Analysis of Psychological Safety and Safety to Experiment

In our theory, we argue that Failure Awards increase psychological safety. We explain that in addition to psychological safety, the feeling of being safe to experiment may also be triggered. First, we predict that Failure Awards—irrespective of type—increase the feeling of being safe to admit failure (PS factor). Second, we argue that innovation-type Failure Awards (additionally) trigger the feeling of being safe to experiment and take risks (safety to experiment factor). We use questions from the post-experimental questionnaire and apply principal component analysis to extract factors based on these two items, i.e., participants' perception of *feeling safe to admit failures* (PS factor, Table 5, Panel A) and their perception of *feeling safe to take risks and experiment* (safety to experiment factor, Table 5, Panel B).

[Insert Table 5 about here]

The PS factor (safety to experiment factor) has an eigenvalue of 2.11 (1.88) and a Kaiser–Meyer–Olkin (KMO) measure of 0.62 (0.502).[18] Using pairwise comparisons (Table 6, Panel A), we find that participants feel significantly safer to admit failure (PS factor) when Failure Awards are present rather than absent ($t = 6.26$, $p < 0.01$). We find this effect for both Failure Award types (innovation-type: $t = -6.03$, $p < 0.01$ and discontinuation-type: $t = -4.51$, $p < 0.01$).[19] As expected, we do not find a significant difference for the PS factor between the two Failure Award types ($t = -1.33$, $p = 0.186$). These findings therefore support our theory that Failure Awards—irrespective of their type—induce psychological safety, making individuals feel safe to admit failure.

For the safety to experiment factor (Table 6, Panel B), we find that participants in the innovation-type treatment feel significantly safer to take risks and experiment compared to participants in the Failure Award absent condition ($t = -2.06$, $p = 0.041$) and compared to the discontinuation-type condition ($t = -1.71$, $p = 0.089$). We do not find that participants in the discontinuation-type treatment feel significantly safer to take risks compared to the Failure Award absent treatment ($t = -0.20$, $p = 0.839$). Hence, in line with our prediction, the effect of feeling safe to take risks and experiment is only triggered in the innovation-type treatment. These findings explain why we find an EoC-reducing effect for the discontinuation-type Failure Award (H2) but not for the innovation-type Failure Award (RQ2).

For RQ1, we do not find that the PS factor and the safety to experiment factor are

[18] A minimum KMO value of 0.5 is necessary for reliable factor estimation (Kaiser, 1970).

[19] These and the following results for the PEQ items are based on the full sample ($n = 264$), as all participants were exposed to the Failure Award manipulation before selecting the investment project.

additive, resulting in a statistically insignificant difference in risk-taking between the two Failure Award types (Table 3, Panel B). To further explore this finding, we extract a third factor that—in contrast to the safety to experiment 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 investment decision.[20] This allows us to examine how participants perceived their environment at the time risk-taking was measured. This third factor has an eigenvalue of 1.81 and an overall KMO of 0.57 (Table 7). Contrary to the safety to experiment 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 (but not at the end) of the experiment, the discontinuation-type Failure Award has a similar effect on participants' risk perception compared to the innovation-type Failure Award. This supports our argument that by communicating that all consequences of risk-taking, particularly failure, are acceptable, the firm does the utmost to increase risk-taking. However, any additional emphasis on risk-taking through the innovation type does not yield an extra effect.

Overall, consistent with our predictions, we find that the feeling of being safe to admit failures (PS factor) is positively linked to risk-taking ($t = 3.80$, $p < 0.01$, untabulated) and negatively related to EoC ($t = -2.36$, $p = 0.019$, untabulated). In contrast, the feeling of being safe to experiment and take risks (safety to experiment factor) is positively linked to EoC ($t = 10.65$, $p < 0.01$, untabulated).

[Insert Table 6 and 7 about here]

Sunk Costs and Escalation of Commitment

Brockner et al. (1981) show that sunk costs may influence escalation behavior. We use the following item from Brockner et al. (1981) to measure the relevance of sunk costs: *"I had*

[20] In contrast, the safety to experiment factor was measured after the EoC task in the PEQ at the end of the experiment.

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 the Failure Award present treatments is 3.33 and 4.09 for the Failure Award absent treatment. The difference is statistically significant ($t = -3.35$, $p < 0.01$). It indicates that Failure Awards decrease participants' sensitivity to sunk costs. Furthermore, we find that 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 termination is supported by the organization, which decreases the perceived need to explain why organizational resources were wasted (Arkes & Blumer, 1985; Sleesman et al., 2012). Thus, Failure Awards indirectly decrease escalation tendencies by reducing decision-makers' sensitivity to sunk costs.

The Impact of Failure Awards on Delayed Project Termination

Failure Awards may induce the feeling of being safe to take risks and experiment (safety to experiment factor), which could lead to a delayed termination decision instead of immediate termination. Hence, our experimental design incorporates a second decision round in which participants again receive a project update indicating a lower expected return after they decide to continue the already poorly performing project.

ANOVA results show no difference among the three treatments concerning delayed EoC ($F = 0.540$, $p = 0.727$, untabulated). Using pairwise comparisons, none of the three comparisons significantly differ 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 when decision-makers delay their discontinuation decision ($t = -0.28$, $p = 0.779$, untabulated).

V. Conclusion

Employees' "fear of failure" may harm firm profitability. On the one hand, it increases employees' reluctance to take sufficient risks, e.g., when starting innovative but risky projects.

On the other hand, it leads to escalation of commitment, i.e., the tendency to overinvest in failing projects (Staw, 1976). To counteract these issues, a growing number of firms have started to grant Failure Awards, rewarding employees who started risky but economically preferred projects that ultimately failed. In this study, we examine the effectiveness of Failure Awards for increasing risk-taking and reducing EoC. We investigate two types of Failure Awards, those that emphasize innovation and risk-taking (innovation-type Failure Awards) and those that concern the early termination of failing projects (discontinuation-type Failure Awards).

We have conducted an online experiment on MTurk in which participants first decided whether to invest in a riskier but economically preferred project or a safer project (our proxy for risk-taking). Next, they had to determine whether to terminate or continue the project when failure became imminent (our proxy for EoC).

We provide evidence that both the presence of Failure Awards and their type affect risk-taking and EoC. Specifically, we find that risk-taking is encouraged through both types of the Failure Award. Furthermore, we find that discontinuation-type Failure Awards decrease EoC. However, we do not find this de-escalating effect for innovation-type Failure Awards.

Moreover, we have predicted and shown that Failure Awards induce psychological safety, which mitigates the fear of failure. Psychological safety creates a feeling of being safe to admit failure. However, through a Failure Award, decision-makers may also feel safe to experiment and take risks. The latter effect is exclusively triggered by innovation-type Failure Awards. Whereas risk-taking is encouraged through psychological safety but not further affected by the feeling of being safe to experiment, the effects on EoC are opposing. Thus, when feeling safe to experiment and take risks, individuals with an innovation-type Failure Award take the risk of further investing in a failing project, which reduces the de-escalating effect of the Failure Award.

Accordingly, our findings have important implications for the design of management control systems. First, they 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 way to reduce EoC. Our results imply that it is crucial for firms to pay close attention to the specific aspects Failure Awards highlight. A focus on innovation and risk-taking, 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 the driving factor of the effect of Failure Awards on risk-taking and escalation behavior. Referring to Barton et al. (1989), who donot find a decrease in EoC when employing an open error management climate, discontinuation-type Failure Awards seem to overcome this challenge by incentivizing project discontinuation.

Future research could further explore this field of research. While we associate Failure Awards with a small monetary reward, future research should investigate whether our results hold when completely non-monetary Failure Awards are employed. Even though our design assures that the effectiveness of Failure Awards is not driven by the (symbolic) monetary component, it would be interesting to see whether granting a trophy or applause has stronger effects within a non-online setting, e.g., a laboratory. Additionally, whereas in our setting the early termination of a failing project is rational and thus preferred, one could investigate whether Failure Awards could lead to the irrational early termination 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 therefore examine the effects of Failure Awards that highlight both aspects equally. Finally, it could be investigated whether Failure Awards, as a potential de-escalation tool, also combat other biases. Since Failure Awards turn mistakes into “something less negative”, an individual’s overly optimistic self-assessment, also known as the *overconfidence bias* (Moore & Healy, 2008), might be attenuated.

References

Anthony, S. D. (2020). *Eat, Sleep, Innovate: How to Make Creativity an Everyday Habit Inside Your Organization*. Harvard Business Review Press. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6026196.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140. https://doi.org/10.1016/0749-5978(85)90049-4.

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64, Part 1*(6), 359–372. https://doi.org/10.1037/h0043445.

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*(1), 45–68. https://doi.org/10.1002/job.179.

Barki, H., Rivard, S., & Talbot, J. (1993). Toward an Assessment of Software Development Risk. *Journal of Management Information Systems*, *10*(2), 203–225. http://www.jstor.org/stable/40398061.

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*(3), 532–544. https://doi.org/10.1111/j.1540-5915.1989.tb01565.x

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*(1), 205–214. https://doi.org/10.5465/256308.

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*(2), 361–374. https://doi.org/10.1111/jpim.12100.

Bentley, J. W. (2021). Improving the Statistical Power and Reliability of Research Using

Amazon Mechanical Turk. *Accounting Horizons*, Article HORIZONS-18-052. Advance online publication. https://doi.org/10.2308/HORIZONS-18-052.

Birkinshaw, J., & Haas, M. (2016a). Fail more, fail better. *London Business School Review*, *27*(2), 34–37. https://doi.org/10.1111/2057-1615.12116.

Birkinshaw, J., & Haas, M. (2016b). Increase Your Return on Failure. *Harvard Business Review*, 88–93. https://www.qualitycharters.org/wp-content/uploads/2016/05/Increase-Your-Return-on-Failure.pdf.

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*(5), 1345–1362. https://doi.org/10.2139/ssrn.2487396.

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 Publishing—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. https://doi.org/10.1016/j.mar.2020.100691.

Brockner, J. (1992). The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress. *The Academy of Management Review*, *17*(1), 39–61. https://doi.org/10.2307/258647.

Brockner, J., Rubin, J. Z., & Lang, E. (1981). Face-saving and entrapment. *Journal of Experimental Social Psychology*, *17*(1), 68–79. https://doi.org/10.1016/0022-1031(81)90007-X.

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*(1), 3–5. https://doi.org/10.1177/1745691610393980.

Cannon, M. D., & Edmondson, A. C. (2005). Failing to Learn and Learning to Fail (Intelligently). *Long Range Planning*, *38*(3), 299–319. https://doi.org/10.1016/j.lrp.2005.04.005.

Cardinaels, E., Dierynck, B., & Hu, W. (2021). Rejections, Incentives, and Employee Creativity: When Chocolate Is Better Than Cash. *Working Paper, Tilburg University and Georgia Institute of Technology.* Advance online publication. https://doi.org/10.2139/ssrn.3022001.

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. https://doi.org/10.1016/j.jbef.2015.12.001.

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*(1), 63–85. https://doi.org/10.2308/bria.2003.15.1.63.

Choi, J., & Presslee, A. (2023). When and why tangible rewards can motivate greater effort than cash rewards: An analysis of four attribute differences. *Accounting Organizations and Society*, *104*, 101389. https://doi.org/10.1016/j.aos.2022.101389.

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.

Cronin, M., Erkens, D. H., Schloetzer, J. D., & Tinsley, C. H. (2021). How Controlling Failure Perceptions Affects Performance: Evidence from a Field Experiment. *The Accounting Review*, *96*(2), 205–230. https://doi.org/10.2308/TAR-2018-0146.

Crosetto, P., & Filippin, A. (2013). The “bomb” risk elicitation task. *Journal of Risk and Uncertainty*, *47*(1), 31–65. https://doi.org/10.1007/s11166-013-9170-z.

Denison, C. A. (2009). Real Options and Escalation of Commitment: A Behavioral Analysis of Capital Investment Decisions. *The Accounting Review*, *84*(1), 133–155. https://doi.org/10.2308/accr.2009.84.1.133.

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,* 2004. https://www.fastcompany.com/51510/gores-text-innovation.

Edmondson, A. C. (1999). Psychological safety and learning behavior in teams. *Administrative Science Quarterly*, *44*(2), 350–382. https://doi.org/10.2307/2666999.

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). John Wiley & Sons Ltd. https://doi.org/10.1002/9780470696712.ch13.

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*(1), 23–43. https://doi.org/10.1146/annurev-orgpsych-031413-091305.

Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *The Academy of Management Review*, *14*(1), 57–74. https://doi.org/10.2307/258191.

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*(1), 93–114. https://doi.org/10.2308/accr-51447.

Festinger, L. (1957). *A theory of cognitive dissonance*. 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*(3), 428–453. https://doi.org/10.1111/apps.12129.

Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological Safety: A Meta-Analytic Review and Extension. *Personnel Psychology*, *70*(1), 113–165. https://doi.org/10.1111/peps.12183.

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*(5), 1094–1104. https://doi.org/10.1016/j.jbusres.2014.10.012.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Doubleday anchor books. Gold, A., Gronewold, U., & Salterio, S. E. (2014). Error Management in Audit Firms: Error Climate, Type, and Originator. *The Accounting Review*, *89*(1), 303–330. https://doi.org/10.2308/accr-50592.

Gomila, R. (2021). Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis. *Journal of Experimental Psychology. General*, *150*(4), 700–709. https://doi.org/10.1037/xge0000920.

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*(3), 213–224. https://doi.org/10.1002/bdm.1753.

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

Heng, C.-S., Tan, B. C., & Wei, K.-K. (2003). De-escalation of commitment in software projects: Who matters? What matters? *Information & Management*, *41*(1), 99–110. https://doi.org/10.1016/S0378-7206(03)00030-2.

Heninger, W. G., Smith, S. D., & Wood, D. A. (2019). Reward type and performance: An examination of organizational wellness programs. *Management Accounting Research*, *44*, 1–11. https://doi.org/10.1016/j.mar.2019.02.001.

Highhouse, S., & Yüce, P. (1996). Perspectives, Perceptions, and Risk-Taking Behavior. *Organizational Behavior and Human Decision Processes*, *65*(2), 159–167. https://doi.org/10.1006/obhd.1996.0014.

Hirshleifer, D. (1993). Managerial Reputation and Corporate Investment Decisions. *Financial Management*, *22*(2), 145. https://doi.org/10.2307/3665866.

Hunt, N. C., & Scheetz, A. M. (2019). Using MTurk to Distribute a Survey or Experiment: Methodological Considerations. *Journal of Information Systems*, *33*(1), 43–65. https://doi.org/10.2308/isys-52021.

Huo, K. (2020). Performance Incentives, Divergent Thinking Training, and Creative Problem Solving. *Journal of Management Accounting Research*, *32*(1), 159–176. https://doi.org/10.2308/jmar-52479.

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*(2), 229–254. https://doi.org/10.1111/j.1744-6570.1977.tb02091.x.

Jeffrey, S. A. (2009). Justifiability and the Motivational Power of Tangible Noncash Incentives. *Human Performance*, *22*(2), 143–155. https://doi.org/10.1080/08959280902743659.

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*(4), 1411–1434. https://doi.org/10.1111/1911-3846.12147.

Kahn, W. A. (1990). Psychological conditions of personal engagement and disengagement at work. *Academy of Management Journal*, *33*(4), 692–724. https://doi.org/10.2307/256287.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, *47*(2), 263. https://doi.org/10.2307/1914185.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, *35*(4), 401–415. https://doi.org/10.1007/BF02291817.

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*, *24*(2), 299–325. https://doi.org/10.2307/3250940.

Kelly, K., Presslee, A., & Webb, R. A. (2017). The Effects of Tangible Rewards versus Cash Rewards in Consecutive Sales Tournaments: A Field Experiment. *The Accounting Review*, *92*(6), 165–185. https://doi.org/10.2308/accr-51709.

Knudsen, M. P., Zedtwitz, M. von, Griffin, A., & Barczak, G. (2023). Best practices in new product development and innovation: Results from PDMA 's 2021 global survey. *Journal of Product Innovation Management*, *40*(3), 257–275. https://doi.org/10.1111/jpim.12663.

Kreilkamp, N., Matanovic, S., Sommer, F., & Wöhrmann, A. (2021). The Effect of Compensation Caps on Risk-Taking. *Journal of Management Accounting Research*, *33*(3), 77–95. https://doi.org/10.2308/JMAR-18-053.

Kuvalekar, A., & Ravi, N. (2019). When and How to Reward Bad News. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.3281644.

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.

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*(6), 291–310. https://doi.org/10.2308/tar-2017-0284.

Lourenço, S. M. (2016). Monetary Incentives, Feedback, and Recognition—Complements or Substitutes? Evidence from a Field Experiment in a Retail Services Company. *The Accounting Review*, *91*(1), 279–297. https://doi.org/10.2308/accr-51148.

Mahlendorf, M. D. (2015). Allowance for failure: reducing dysfunctional behavior by innovating accountability practices. *Journal of Management & Governance*, *19*(3), 655–686. https://doi.org/10.1007/s10997-013-9276-3.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review*, *115*(2), 502–517. https://doi.org/10.1037/0033-295X.115.2.502.

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.

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*(3), 259–269. https://doi.org/10.1002/cjas.219.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running Experiments Using Amazon Mechanical Turk. *Judgment and Decision Making*, *5*(5), 411–419. https://doi.org/10.1017/S1930297500002205.

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. https://doi.org/10.1016/j.jesp.2017.01.006.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods*, *46*(4), 1023–1031. https://doi.org/10.3758/s13428-013-0434-y.

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*(5-6), 757–786. https://doi.org/10.1007/s11573-020-00987-x.

Schoemaker, P. J. H. (1982). The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations. *Journal of Economic Literature*, *20*(2), 529–563. http://www.jstor.org/stable/2724488.

Schoorman, F. D., & Holahan, P. J. (1996). Psychological antecedents of escalation behavior: Effects of choice, responsibility, and decision consequences. *Journal of Applied Psychology*, *81*(6), 786–794. https://doi.org/10.1037/0021-9010.81.6.786.

Seybert, N. (2010). R&D Capitalization and Reputation-Driven Real Earnings Management (Partially Retracted). *The Accounting Review*, *85*(2), 671–693. https://doi.org/10.2308/accr.2010.85.2.671.

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*(4), 419–426. https://doi.org/10.1037/0021-9010.77.4.419.

Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the Determinants of Risk Behavior. *Academy of Management Review*, *17*(1), 9. https://doi.org/10.2307/258646.

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*(6), 1573–1592. https://doi.org/10.2307/256844.

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*(3), 541–562. https://doi.org/10.5465/amj.2010.0696.

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. https://doi.org/10.1016/j.aos.2007.11.001.

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*(1), 27–44. https://doi.org/10.1016/0030-5073(76)90005-2.

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*(4927), 216–220. https://doi.org/10.1126/science.246.4927.216.

Steinkühler, D., Mahlendorf, M. D., & Brettel, M. (2014). How Self-Justification Indirectly Drives Escalation of Commitment. *Schmalenbach Business Review*, *66*(2), 191–222. https://doi.org/10.1007/BF03396905.

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/.

Tafkov, I. D. (2013). Private and Public Relative Performance Information under Different Compensation Contracts. *The Accounting Review*, *88*(1), 327–350. https://doi.org/10.2308/accr-50292.

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

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*(6), 1228–1240. https://doi.org/10.1037/0021-9010.90.6.1228.

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

Wang, L. W. (2017). Recognizing the Best: The Productive and Counterproductive Effects of Relative Performance Recognition. *Contemporary Accounting Research*, *34*(2), 966–990. https://doi.org/10.1111/1911-3846.12292.

Wiseman, R. M., & Gomez-Mejia, L. R. (1998). A Behavioral Agency Model of Managerial Risk Taking. *Academy of Management Review*, *23*(1), 133–153. https://doi.org/10.5465/amr.1998.192967.

Wong, K. F. E. (2005). The Role of Risk in Making Decisions under Escalation Situations. *Applied Psychology*, *54*(4), 584–607. https://doi.org/10.1111/j.1464-0597.2005.00236.x.

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*(2), 282–297. https://doi.org/10.1037/0021-9010.91.2.282.

Wu, H.-L. (2008). When does internal governance make firms innovative? *Journal of Business Research*, *61*(2), 141–153. https://doi.org/10.1016/j.jbusres.2007.06.010.

Zhou, J., & George, J. M. (2001). When Job Dissatisfaction Leads to Creativity: Encouraging the Expression of Voice. *Academy of Management Journal*, *44*(4), 682–696. https://doi.org/10.5465/3069410.

TABLE 1

Payoff table

	failing project continued		**failing project terminated (and alternative investment realized)**	
			Failure Award present	Failure Award absent
Initial Investment	3 m lira	3 m lira	3 m lira	3 m lira
Additional Investment	1 m lira	1 m lira	1 m lira	1 m lira
	Best-Case Scenario	*Worst-Case Scenario*		
Probability	33%	67%	100%	100%
Expected cash inflows (updated)	7 m lira	3 m lira	4.8 m lira	5 m lira
Expected Project Value	**0.32 m lira**		**0.8 m lira**	**1 m lira**
Initial Project Balance	5 m lira	5 m lira	5 m lira	5 m lira
Expected Balance on Project Account	**5.32 m lira**		**5.8 m lira**	**6 m lira**
Performance-contingent compensation (1% of expected value)	53,200 lira		58,000 lira	60,000 lira
Failure Award			2,000 lira	
Total expected compensation	**53,200 lira**		**60,000 lira**	**60,000 lira**

TABLE 2

Descriptive statistics (mean, [standard deviation])

	Failure Award Present[a]			Failure Award absent	**Total**
	Innovation-type	Discontinuation-type	**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 Failure Award* is manipulated between subjects on two levels. In the innovation-type treatment, participants were 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 were 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 percentage of participants who chose the risky project *Smart Mop Robot*.

[c] *Willingness of project continuation* represents the indicated percentage (on a 101-point scale with 0% = termination and 100% = continuation) to which participants were willing to continue the failing 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 invested in the riskier project. Thus, Panel B contains only the results for these participants (n = 165).

TABLE 3

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: 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 [RQ1]	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. The safe project is the *Smart Vacuum Robot* project.
[b] The factor *Treatments* has three levels: 1) Discontinuation-type, 2) Innovation-type and 3) No Failure Award.
[c] All p-values are two-tailed.

TABLE 4

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 three groups: 1) Discontinuation type, 2) Innovation type and 3) No Failure Award.

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

TABLE 5

Factor Analyses: PS Factor and Safety to Experiment Factor

Panel A: Psychological Safety – PS Factor

Questions (7-point scale)

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 Experiment Factor

Questions (7-point scale)

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 6

Factor Analysis on PS Factor and Safety to Experiment Factor

“Psychological Safety – PS Factor”

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

“Safety to Experiment Factor”

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 7

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)

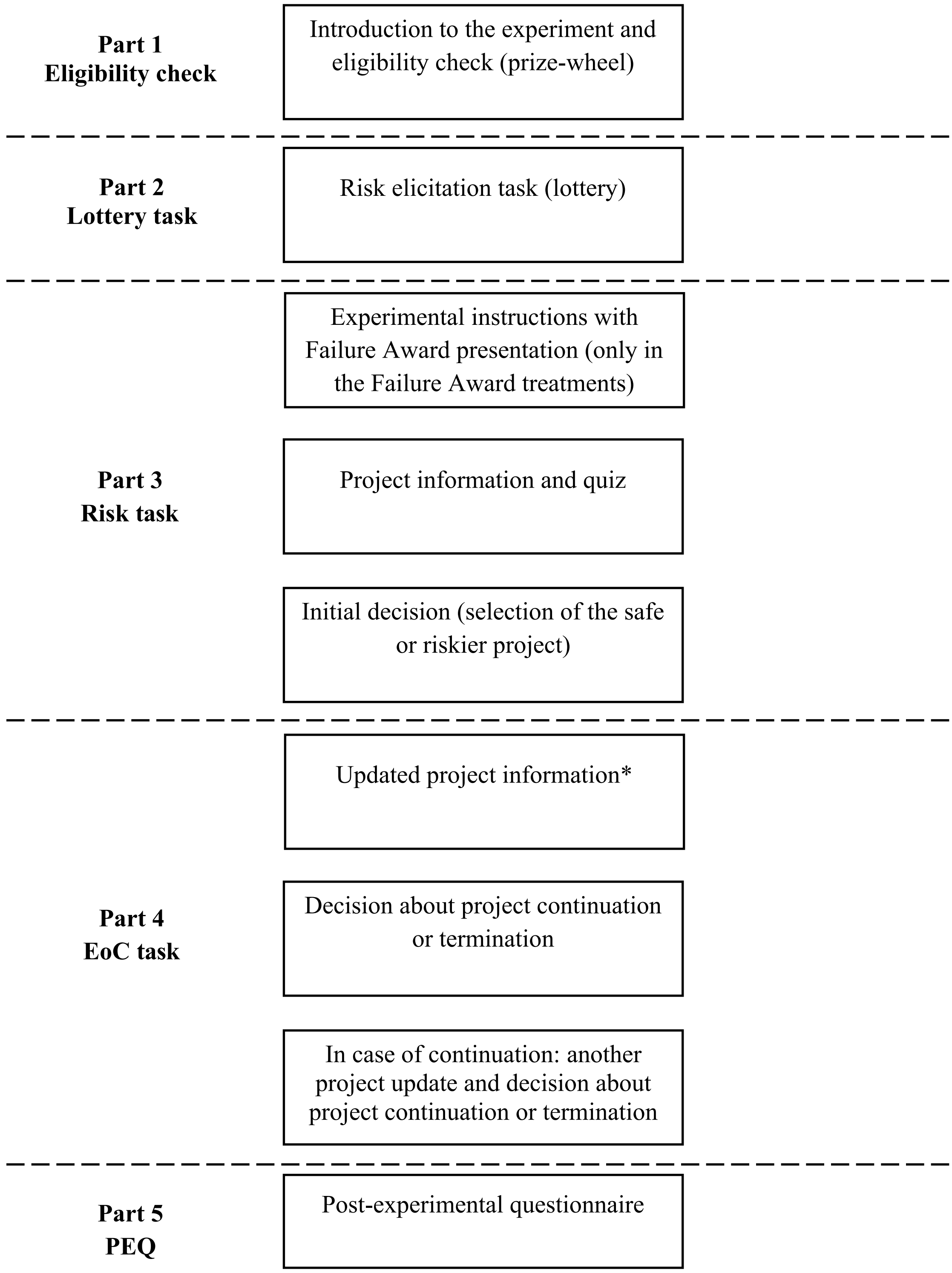
2. How would you characterize your selected project?
(endpoints: potential for loss and potential for gain)

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 who chose 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?	***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**.	*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 **take the risk to start a highly innovative 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 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.	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

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.

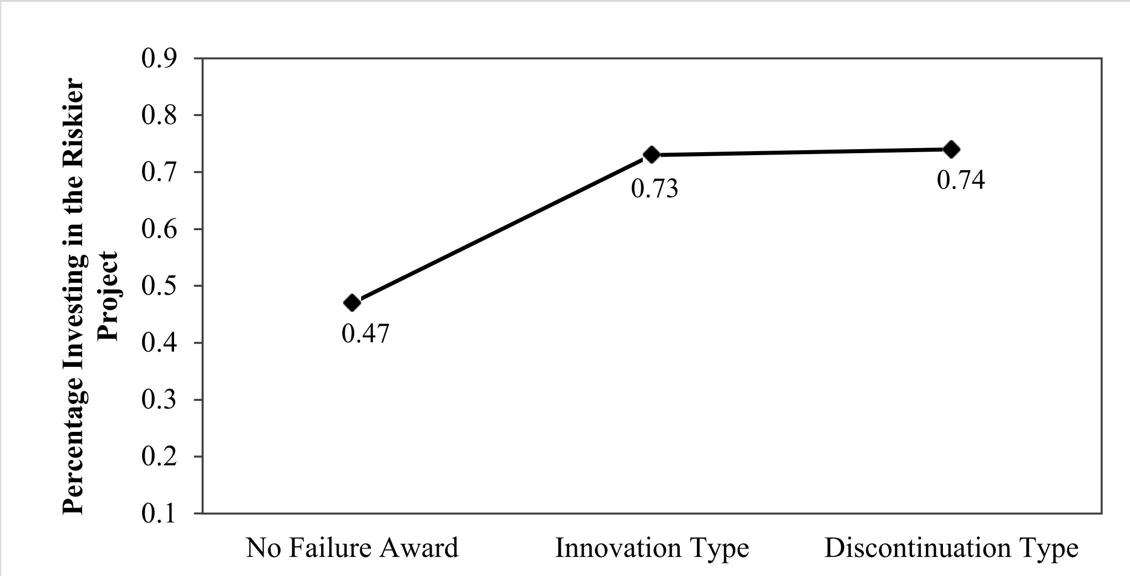
The following **predicted values** were calculated based on currently known information:

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

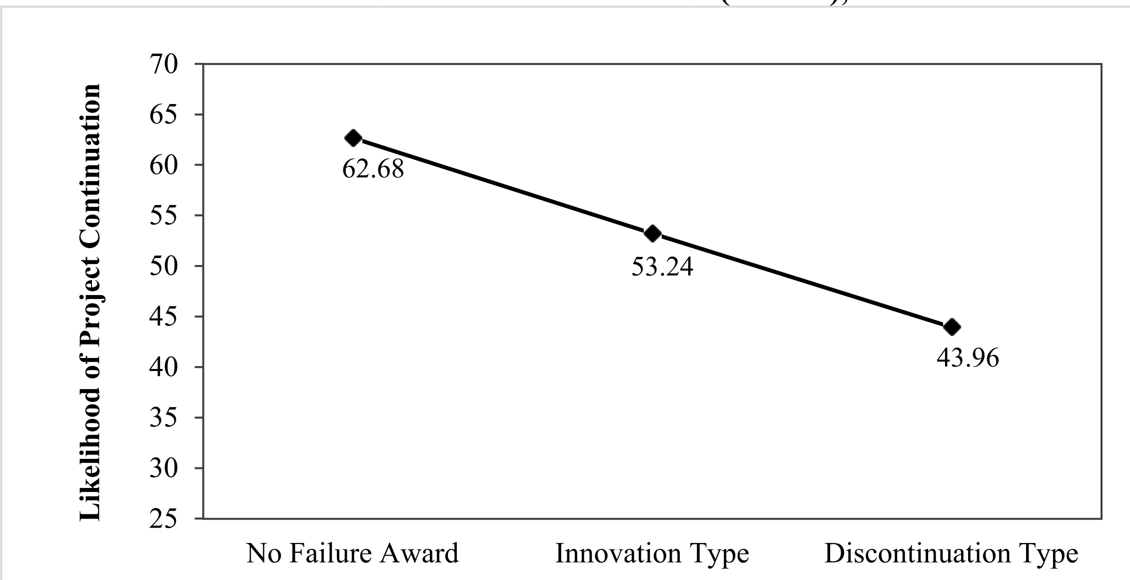
FIGURE 5

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

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



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



[a] 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 Failure Award on two levels (innovation vs. discontinuation) and added a Failure Award absent treatment.

<table>
<tr><td>Study no.</td><td>2</td></tr>
<tr><td>Title</td><td>The Effect of Monitoring on Teleworkers’ and Office Workers’ Behavior</td></tr>
<tr><td>Authors</td><td>Rebecca Sabel, Niklas Kahl, Arnt Wöhrmann, Corinna Ewelt-Knauer</td></tr>
<tr><td>Author contribution</td><td>
<table>
<tr><td></td><td>Sabel</td><td>Kahl</td><td>Wöhrmann</td><td>Ewelt-Knauer</td></tr>
<tr><td>Numeric share</td><td>0.5</td><td>0.4</td><td>0.05</td><td>0.05</td></tr>
<tr><td>Conceptual development of research question</td><td>✓</td><td></td><td>✓</td><td>✓</td></tr>
<tr><td>Development of theory</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>Methodology</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>Acquisition of data</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>Analysis/interpretation of data</td><td>✓</td><td>✓</td><td></td><td></td></tr>
<tr><td>Writing the manuscript</td><td>✓</td><td>✓</td><td>✓</td><td></td></tr>
</table>
</td></tr>
<tr><td>Publication status</td><td>Under review
Journal of Management Accounting Research (VHB-Rating 2024: A)</td></tr>
<tr><td>Peer-reviewed conferences</td><td>
<table>
<tr><td>2025</td><td>American Accounting Association (AAA) Annual Meeting 2025 (Chicago) – Accepted, presentation forthcoming (August 2025)</td></tr>
<tr><td>2025</td><td>European Institute For Advanced Studies In Management (EIASM) Conference 2025 – Submitted for presentation (September 2025)</td></tr>
</table>
</td></tr>
<tr><td>Research approach</td><td>Experimental study</td></tr>
<tr><td>Language</td><td>English</td></tr>
<tr><td>Abstract</td><td>Today’s working environment is shaped by two megatrends: telecommuting and surveillance. While both are widespread in practice, research on how telecommuting affects employee behavior and interacts with monitoring remains scarce. In an experiment, we examine differences in effort and misreporting between teleworkers and office workers. We manipulate the presence of monitoring and focus on a setting that allows employees to reciprocate or retaliate against their employer. Consistent with our predictions, teleworkers exhibit greater effort and misreport less than office workers. A path analysis reveals that these effects are driven by reciprocity. Further, we find that monitoring leads to a greater reduction in effort and misreporting among office workers compared to teleworkers. Our study provides important implications for the design and implementation of management control systems.</td></tr>
</table>

The Effect of Monitoring on Teleworkers' and Office Workers' Behavior

Rebecca Sabel
Justus Liebig University Giessen (Germany)
Rebecca.Sabel@wirtschaft.uni-giessen.de
Licher Str. 62,
35394 Giessen, Germany

Niklas Kahl
Justus Liebig University Giessen (Germany)
Niklas.Kahl@uni-giessen.de

Arnt Wöhrmann
Justus Liebig University Giessen (Germany)
Arnt.Woehrmann@wirtschaft.uni-giessen.de

Corinna Ewelt-Knauer
Justus Liebig University Giessen (Germany)
Corinna.Ewelt-Knauer@wirtschaft.uni-giessen.de

Abstract

Today's working environment is shaped by two megatrends: telecommuting and surveillance. While both are widespread in practice, research on how telecommuting affects employee behavior and interacts with monitoring remains scarce. In an experiment, we examine differences in effort and misreporting between teleworkers and office workers. We manipulate the presence of monitoring and focus on a setting that allows employees to reciprocate or retaliate against their employer. Consistent with our predictions, teleworkers exhibit greater effort and misreport less than office workers. A path analysis reveals that these effects are driven by reciprocity. Further, we find that monitoring leads to a greater reduction in effort and misreporting among office workers compared to teleworkers. Our study provides important implications for the design and implementation of management control systems.

Keywords: teleworking; monitoring; effort; misreporting; reciprocity

JEL Classifications: M41; M50; J81

Data Availability: Contact the authors.

We appreciate the financial support provided by the Dr. Werner Jackstädt-Foundation. All authors have no financial conflicts of interest related to this research.

I. Introduction

Telecommuting, i.e., working at least partially away from a central workplace, has become the "new normal" in many firms – boosted particularly by the COVID-19 pandemic (Allen et al. 2015).[21] Recently, however, more and more companies have returned to or at least have articulated strong preferences for office-based work – mainly because of fears of shirking and dishonest behavior by teleworkers (Business Insider 2024a). At the same time, employee surveillance has spread in the business world (BBC 2020; Waterson 2016). Firms that have committed to telecommuting now not only apply this technology in the office but also in telecommuting environments. In this study, we first examine whether the widespread assumption that telecommuters work less and are more dishonest vis-à-vis office workers is justified. Next, we examine the effects of employee surveillance and discuss whether this practice affects office and teleworkers differently.[22]

Telecommuting and employee surveillance have become two megatrends in the workplace. Concerning the first trend, a 2022 survey reveals that 87% of U.S. employees with remote-work options work at least one day per week outside the office (McKinsey & Company 2022). Similarly, 56% of U.S. respondents express a strong preference for telework (Piacenza et al. 2022). This preference is rooted in the advantages associated with telecommuting, such as reduced commuting time, cost savings, flexibility and better work-life balance (Global Workplace Analytics 2021).

While also firms benefit from telecommuting, e.g., by less office space and estimated annual savings of $11,000 per half-time remote worker (Global Workplace Analytics 2021), many firms such as Amazon (2024) have started to require employees to return to the office. Yet, surveys document that firms' decision to offer only office-based work might be associated with high costs. For example, in case of Amazon, 91% of Amazon employees are dissatisfied

[21] We use the terms telecommuting, teleworking, working from home and remote work interchangeably.

[22] We use the terms monitoring and surveillance interchangeably.

with the decision, and thus 72% have started looking for alternative jobs (Chen 2024). This is in line with other studies, showing that 66% of U.S. employees would quit their jobs when the remote work option is removed and about 46% reveal that they would be less willing to go the extra mile when forced to work in the office (Owl Labs 2022).

Due to the high costs of not offering the option for telecommuting, it is important to better understand employee behavior in the office vis-à-vis when telecommuting. However, research in this field provides mixed evidence. While some researchers find positive effects of telecommuting on productivity (Belanger 1999; Bloom et al. 2015; Rupietta and Beckmann 2018), others document negative (Gibbs et al. 2021; Golden et al. 2008), or even no effects at all (Gajendran and Harrison 2007).[23] Prior research speculates that these inconsistencies are driven by methodological issues, e.g., single-source data, cross-sectional designs, or self-reported variables (e.g., Allen et al. 2015). While we employ a controlled experiment that addresses many of these issues, our study differs from prior research in one important aspect: We acknowledge that employees are well aware of the fact that employers today often *have the option* to allow telecommuting, i.e., telecommuting is an employer's decision. In our experimental study, employees can directly reciprocate or retaliate against their employer for the telecommuting decision. We argue that this feature better reflects the real world and allows for unambiguous predictions.

Firms' reservations concerning telecommuting can be easily explained by the assumptions of agency theory. According to agency theory, employees will exploit the loss of control resulting from telework to maximize their utility (Ross 1973). This leads to a reduction of effort due to increased opportunity costs when not working in the office (e.g., engaging in personal activities) and increased misreporting (e.g., of working hours). In this vein, Elon Musk posted on X that everyone who disagrees with a policy of coming back to the office "should

[23] Another example is Kelly et al. (2024) who, in their survey, do not find any direct effect of telework on perceived productivity but an indirect effect through more hours worked. Both variables are self-reported by MTurk respondents in this study.

pretend to work somewhere else" (Porter 2022). According to agency theory, one option—besides incentives—to address this issue is to increase monitoring (Jensen and Meckling 1976).

Employers' desire to monitor employees is not new and had come up long before teleworking became popular. Technological advances have made employee monitoring cheaper, more accessible and more reliable, resulting in an increased usage (Alge and Hansen 2013; Ravid et al. 2020). Regarding the other workplace trend we address, a recent study finds that 49% of office workers in the U.S. are monitored at work (American Psychological Association 2023). Many of the systems used at the workplace monitor whether the employee is physically present at work, e.g., sitting at the computer or moving the mouse. Often output cannot be reliably measured and if it can, it is more attractive to use incentives instead of monitoring. Capturing physical presence means verifying that the employee is at work. In this respect, the British newspaper Daily Telegraph placed sensors beneath employees' desks using heat and motion detection to monitor whether employees sit at their desks (Waterson 2016). Similarly, the UK head office of Barclays pilot-tested a system tracking employees' computer usage in the office (BBC 2020). This was followed by a controversial initiative where the company secretly placed black boxes under employees' desks to monitor their whereabouts (Ball 2021). Another example is Canon, a global imaging solutions company, which offers workplace AI solutions combining cameras and AI video analytics (Canon 2024). These systems monitor employees in the office in real time.

During the COVID-19 pandemic, the usage of employee monitoring increased and the technology was also used to track teleworkers. Also, the demand for surveillance software rose by 108% in April 2020 compared to the previous year (Ball 2021). Consistent with this trend, employee-surveillance-software providers reported a surge in sales inquiries with increases ranging from 333% (Desk Time software) to 139% (KickIdler software) in 2020 compared to the previous year (Ball 2021; Brown 2020). Importantly, these surveillance technologies are also used for teleworkers and, e.g., make screenshots or record videos using the computer's webcam to verify whether the employee is really sitting in front of the computer. Against this backdrop, our study not only examines office workers' and teleworkers' working behavior, i.e., their effort and misreporting, but also their reactions to increased monitoring.

Our first two hypotheses predict that teleworkers will reciprocate their employer for being allowed to telework by more effort (H1a) and less misreporting (H1b) compared to office workers. As employees today generally have a preference for telecommuting, teleworkers will reciprocate and office workers will retaliate against their employer (Greer and Payne 2014). Our remaining hypotheses predict a moderating effect of monitoring for effort and misreporting. For effort (H2a), we predict that monitoring has a more negative effect on office workers compared to teleworkers. It follows self-determination theory that monitoring crowds out intrinsic motivation via reduced autonomy (Christ et al. 2012; Enzle and Anderson 1993; Gagne and Deci 2005; Schedlinksy et al. 2020). Monitoring also induces self-awareness, prompting individuals to reflect on themselves and compare their behavior with salient standards shaped by the situation (Davidson 2019; Lerner and Tetlock 1999; Wicklund and Duval 1971). When there is no salient and established performance standard—as in our setting where employees receive fixed compensation for the task and no specific goals exists—the aspect becoming salient is the employee's "self" (Carver 2012). Employees then reflect upon their current level of reciprocity driven by motivation. Particularly, the negative motivational effect of office work is magnified, which results in even less effort under monitoring.

For misreporting (H2b), a salient and established standard exists as individuals usually share a preference for honesty (Evans et al. 2001). In this case, particularly office workers become aware that their intention to misreport due to low reciprocity would move them away from the salient standard of honesty. Thus, office workers misreport less under monitoring and the discrepancy with teleworkers (who report more honestly even in the absence of monitoring) is reduced.

To test our predictions, we conduct a 2×2 between-subjects experiment, manipulating monitoring (present vs. absent) and location (office vs. telework).[24] Monitoring is implemented by video and audio monitoring. In terms of location, participants either complete the experiment in the laboratory (office) or from home (telework). Before the main experiment takes place, a small group of participants assume the role of an employer and decide where their (assigned) employees should work and whether they should be monitored. During the main experiment, other participants assume the role of their employees. Employees work on two tasks: a slider task and a reporting task. We measure effort by the number of sliders moved. While employee-participants receive a fixed compensation, their employer's compensation increases with the number of sliders solved. We use the reporting task to measure misreporting.While misreporting during this task increases employee's payoff, it decreases the employers' compensation.

In line with our hypotheses, we find that teleworkers exhibit greater effort (H1a) and misreport less (H1b) compared to office workers. When monitoring is present, effort and misreporting decrease more among office workers than among teleworkers. Thus, H2a and H2b are supported. Moreover, additional analyses reveal that reciprocity mediates the relationship between location (i.e., telework or office) and effort as well as misreporting. Further, we show that monitoring induces self-awareness and thus has an impact on how reciprocity (retaliation) translates into lower effort and misreporting.

Our study contributes to both theory and practice. First, from a theory perspective, we show that teleworkers reciprocate their employer for the opportunity to work remotely. Our study differs from prior research, as employees in our experiment are aware of the fact that it is their employer's decision to allow telework or to force them to work in the office. As employers' compensation is linked to employees' working behavior, reciprocity can unfold.

[24] The research was conducted in an ethical manner. Specifically, subjects were treated anonymously in accordance with the relevant data protection regulations and were not exposed to specific risks. Furthermore, subjects were not deceived in any way or at any time. The institution at which the study was conducted does not have a review board to provide ethical clearance.

Hence, we contribute by providing a new perspective on the results of prior research (Bloom et al. 2015; Brüggen et al. 2024; Gajendran and Harrison 2007; Gibbs et al. 2021; Golden et al. 2008). For example, Brüggen et al. 2024 examine a setting that is very similar to ours and do not observe significant differences in effort and misreporting between randomly assigned office and teleworkers.[25] Yet, this study differs from ours because—among other things—these authors are interested in a setting where reciprocity does not matter and is, thus, unlikely to affect behavior. Taken together, our study makes practitioners who worry about the negative effects of telework due to loss of control aware of the positive effects of telecommuting resulting from reciprocity.

Second, to the best of our knowledge, this study is the first to examine whether the effects of monitoring depend on location, i.e., whether employees are monitored in the office or while telecommuting.[26] Answering this research question is important as employee surveillance has become widespread (Fortune Business Insights 2024) and is no longer used only in the office but also applied to teleworkers (European Digital Rights 2022). Notably, we examine this interaction effect not only for effort but also for misreporting. As monitoring is argued to lower intrinsic motivation it is important to investigate work dimensions that either depend on intrinsic motivation (i.e., effort) or do not (i.e., misreporting). Investigating both dimensions in a single study helps to provide potential explanations for the mixed findings of monitoring on effort and honesty in prior research (Beck et al. 2018; Dickinson and Villeval 2008; Kroher and Wolbring 2015).

[25] Brüggen et al. (2024) measure effort based on the number of sliders correctly solved. In contrast we operationalize effort through the number of sliders moved.

[26] Ko and Baek (2024) use a field-study and implement computer monitoring in addition to a set of output-controls. However, only teleworkers were monitored and their finding—that monitored teleworkers exhibit a higher productivity level compared to unmonitored office workers—is based on the absence of an ideal control group. Hence, using an experiment and holding the treatments constant allows us to examine whether monitoring effects depend on the location. Moreover, Kelly et al. (2024) find in their survey that more intensive monitoring is generally associated with higher perceived productivity. However, when employees work remotely, this relationship reverses: in remote settings, more intensive monitoring tends to be associated with lower perceived productivity. It is important to note that these findings are based on self- reports, and no direct comparison to office workers is provided.

Third, our study examines a setting with imperfect monitoring. In other words, dysfunctional behavior cannot be effectively observed by the employer and, thus, cannot be sanctioned. For this reason, our findings are more generalizable to practice, where perfect monitoring is often too expensive or not feasible. From a theory perspective, this provides clean evidence for the effects of monitoring that are not mixed with the effect of (different types/levels of) sanctions.

II. Research Setting and Hypotheses

Research Setting

The research setting we examine can be characterized by three elements: First, we incorporate employer's deliberate decisions to allow or deny telecommuting and to implement monitoring or not. Employees can react to this decision by reciprocating (or retaliating) to their employer. Second, we implement employee surveillance as commonly observed in practice and capture employees' physical presence. We do so because output is often not available and if it is available other management controls such as incentives can be used. Third, we use imperfect monitoring, i.e., dysfunctional activities are not perfectly detected by the employer. We believe that imperfect monitoring better reflects the real world where employees use mouse jigglers etc. to pretend keyboard activity, for example. Perfect monitoring—if feasible at all—would be expensive for firms but of course could also lead to a first-best-solution.

Hypotheses Development

Location

Teleworking can be generally defined as "[…] a work practice that involves members of an organization substituting a portion of their typical work hours (ranging from a few hours per week to nearly full-time) to work away from a central workplace—typically from home—using technology to interact with others as needed to conduct work tasks" (Allen et al. 2015, 44).[27] Today, employees have—on average—a strong preference for telework (Da Silva et al. 2023; Piacenza et al. 2022). This becomes particularly obvious when firms remove the possibility to work from home. A case in point is Deutsche Bank, Germany's largest financial institution. When the bank announced a limit for remote work of 40% of the working hours, it faced intense criticism on its internal platform in response to the announcement and ultimately decided to disable the comment function (Business Insider 2024b; Osman 2024). Many other companies such as Amazon, Meta, and Tesla have also implemented return-to-office mandates, and have faced similar employee resistance, including petitions and protests (Business Insider 2024a).

Based on employees' preference for telework, our first hypothesis predicts that employees will reciprocate the opportunity to work from home. Effort depends on (intrinsic) motivation which is driven—among other factors (e.g., incentives)—by reciprocity. Reciprocity (Gouldner 1960) is a key concept of gift exchange theory (Akerlof 1982). It argues that individuals respond to kind actions with kindness and unkind actions with retaliation (Falk and Fischenbacher 2006). In this vein, employers who allow telework provide a "gift" that employees are likely to reciprocate. Hence, employees show gratitude for the benefits of working from home (e.g., increased flexibility) which increases their motivation, and ultimately also effort. Vice versa, employees who have to work in the office retaliate against their employer.

Various studies—but not all—support the assumption that teleworking positively affects effort. The studies that document a positive effect use diverse methodologies and datasets. For example, Rupietta and Beckmann (2018) employ data from the Socio-Economic Panel and find that working from home increases effort. Similarly, Bloom et al. (2015) observe improved performance among teleworkers in a field experiment with call center agents. The performance

[27] Given this definition, telework encompasses a broad range of working agreements where employees perform their job outside the office. This also includes—but is not limited to—working from home. Therefore, we use both terms interchangeably.

effects are partially attributed to increased effort, as teleworkers spend more minutes actively working per shift. Gibbs et al. (2021) analyze data from an Asian IT services company and find that remote work increased hours worked, i.e., effort, despite a decline in productivity. Similarly, Kelly et al. (2024) find in their survey with MTurk workers that remote work indirectly increases perceived productivity through more hours worked. While these studies suggest that telework increases effort, the results from experimental studies by Dutcher (2012) and Brüggen et al. (2024) provide mixed results. Dutcher (2012) examines the effect of telecommuting on productivity and finds that telecommuting is only beneficial for creative tasks but has—on average—no effect for dull tasks. Brüggen et al. (2024) show that telework has no effect on effort. However, contrary to our study, the latter two experiments do not incorporate the possibility for employees to reciprocate their employer for the decision regarding the working location.[28]

To sum up, we predict that teleworkers who appreciate the benefits of working from home will reciprocate employers allowing telework by increased motivation and, thus, more effort.[29] On the other hand, office workers are likely to respond with resistance or retaliation, leading to lower motivation and, ultimately, reduced effort. We formthat teleworkers work less but also that they hypothesis as follows:

H1a: Effort is higher for teleworkers compared to office workers.

Employers are not only concerned that teleworkers work less but also that they act dis-

---

[28] Participants in Brüggen et al. (2024) earned money either for themselves or for a university fund, supporting education and providing student subsidies. As a result, increased (decreased) effort primarily benefited (harmed) the participants themselves, limiting the scope for genuine reciprocal behavior towards the employer. As our theory focuses on intrinsic motivation we increased the duration of the slider task compared to Brüggen et al. (2024) to make sure that differences in intrinsic motivation between teleworkers and office workers can materialize if they exist.

[29] The argument based on agency theory that teleworkers exploit information asymmetry and work less appears overly simplistic. In the real world, employees working in the office can similarly avoid work (e.g., by chatting with colleagues or browsing the web on their smartphone). This is particularly true when monitoring is absent. The risk of job loss deters opportunistic behavior regardless of location. Instead of focusing solely on the control loss in telework environments, we also consider its benefits—improved work-life balance, satisfaction, reduced commuting, and flexibility (e.g., Aksoy et al. 2023)—which employees reciprocate.

honestly. Another common fear is that teleworkers engage in misreporting and, e.g., exaggerate working hours without accounting for breaks. The U.S. Patent and Trademark Office (USPTO), for example, has been subject to nationwide investigations due to widespread misreporting of working hours (Ko and Baek 2024).

Contrary to the belief that telework leads to more misreporting, we argue—similar to H1a— that reciprocity will instead mitigate it. Offering telework can trigger employees' sense of moral obligation. This in turn encourages teleworkers to repay employers' trust by reporting accurately and truthfully. This is in line with prior literature indicating that individuals repay trust by acting in a trustworthy manner, even when it involves a cost (Camerer 2011; Johnson and Mislin 2011). In contrast, employees required to work from the office perceive their working conditions as unfair and a signal of distrust—particularly when teleworking opportunities are explicitly denied by the employer. Consequently, they respond with negative reciprocity (or retaliation) and report less honestly.

This argument corresponds to the findings of Houser et al. (2012), who observe that individuals feeling treated unfairly are more likely to engage in dishonest behavior to restore the perceived imbalance. Brüggen et al. (2024) find a reducing—but insignificant—effect of teleworking on misreporting when participants are randomly assigned to work locations.[30] They attribute this result primarily to (nine) individuals who preferred working from home but were forced to work at the office, suggesting that perceived unfairness resulted in retaliation (i.e., negative reciprocity).

To sum up, we rely on reciprocity and predict employees report more honestly when their employer deliberately allows telework compared to when the employer requires office attendance. Hence, we formally state our hypothesis as follows:

H1b: Misreporting is lower for teleworkers compared to office workers.

[30] Brüggen et al. (2024) find an increase in misreporting among teleworkers who self-selected into the telework arrangement. The authors conclude that less honest individuals have a preference for telework.

Monitoring

Next, we focus on the moderating effect of monitoring on the behavior of teleworkers vis-à-vis office workers. We argue that monitoring has two effects that we explain first: (a) a crowding-out effect of intrinsic motivation, and (b) a self-awareness effect.

The *crowding-out effect* stems from self-determination theory (SDT) arguing that individuals are inherently self-motivated when basic psychological needs are fulfilled (Ryan and Deci 2000). One of these needs is autonomy, referring to individuals' perceived control over their environments and actions. We argue that a working environment characterized by surveillance leads to lower perceived autonomy. This is in line with prior research showing that external controls can undermine perceived autonomy (Christ et al. 2012; Enzle and Anderson 1993; Schlund and Zitek 2024). SDT further argues that individuals' actions range on a continuum from self-determined and volitional (internal locus of causality) to externally driven by external factors (external locus of causality) (Pelletier et al. 2001; Schedlinksy et al. 2020). As argued by Plant and Ryan (1985), monitoring acts as a mechanism that shifts the locus of causality from internal to external. Monitoring reduces employees' perceived autonomy by diminishing their feelings of self-determination and creating a sense of dependence or lack of control over their actions (Gagne and Deci 2005).[31]

Consequently, we predict that monitoring reduces individuals' perception of autonomy, thereby leading to a decrease in intrinsic motivation. When autonomy is reduced, employees perceive their actions as dictated by external forces rather than their own will. As a result, intrinsic motivation—characterized by engaging in activities out of genuine interest and enjoyment (Farrell et al. 2017; Fessler 2003)—is diminished, resulting in a crowding-out effect of motivation (a).

[31] Noteworthy, SDT does not require an actual shift in the locus of causality for individuals to feel controlled (Deci and Ryan 2012). The perception of control does suffice. Thus, surveillance does not need to impose tangible behavioral restrictions to lower employees' perceived autonomy (Christ et al. 2008; Frey and Jegen 2001; Tessier and Otley 2012). This aligns with research showing that even subtle monitoring cues, such as the mere display of a pair of watching eyes, can significantly alter behavior (Bateson et al. 2006).

In addition, monitoring induces a second effect which is (b) the *self-awareness effect*. Self-awareness is a sub-theory of social facilitation theory and can be defined as "the capacity to focus attention on oneself, and thus to self-evaluate" (Silvia et al. 2004, 475). When individuals become self-aware, they are prompted to evaluate their actions, often striving to attain consistency in their beliefs and behaviors (Davidson 2019; Wicklund and Duval 1971). Situational cues activate mental scripts and lead individuals to become self-aware (Lerner and Tetlock 1999). We follow prior research showing that, e.g., mirrors, the presence of others, or a video camera can induce such self-awareness (Guerin 1993; Plant and Ryan 1985; Wicklund and Duval 1971). In this respect, we argue that monitoring (e.g., via a camera) provides situational cues that induce self-awareness (Miller et al. 2017). We discuss the effects of self-awareness that differ for effort and misreporting below and explain why we refrain from predicting a main effect for monitoring in the next section.

The Moderating Role of Monitoring

Our next hypothesis (H2a) focuses on the moderating role of monitoring on effort. Building on the two effects discussed above, we predict that monitoring leads to a greater decrease in effort for office workers compared to teleworkers. Effort, among other factors, depends on (intrinsic) motivation, particularly when compensation is not linked to effort (e.g., under fixed-pay contracts).

On the one hand, intrinsic motivation stems from the nature of the task when employees find a task engaging and enjoyable (Farrell et al. 2017; Fessler 2003). On the other hand, intrinsic motivation depends on the level of perceived autonomy (Ryan and Deci 2000). Holding the task and, thus, the joy associated with the task constant for office workers and teleworkers, any difference in the level of intrinsic motivation between these two groups of employees is rooted in the level of perceived autonomy. The crowding-out effect of monitoring harms employees' perceived autonomy and results in less intrinsic motivation and effort.[32]

At the same time, monitoring makes employees more self-aware, i.e., they reflect on themselves and compare their behavior with salient standards shaped by the situation (Davidson 2019; Wicklund and Duval 1971). When employees conduct an effort task that depends on intrinsic motivation and is not otherwise enforced (e.g., by incentives), there is no salient or universal standard that the individual can compare his or her behavior to. The level of intrinsic motivation differs between individuals, and a not incentivized, i.e., fixed compensated, task, can be interpreted as a "do your best goal" or even be perceived as a situation where exerting effort is not valued. If there is no standard, the aspect becoming salient under increased self-awareness can be "the self" of the agent and his or her internal stimuli such as emotions, aches or pains (Carver 2012). Carver (2012, 53) concludes "whatever aspect of the self was salient at the moment attention was self-directed would have a disproportionate influence on the person's subsequent subjective experience and behavioral response".

Hence, when employees conduct a task that depends on (intrinsic) motivation, intrinsic motivation becomes salient and self-aware employees reflect more on their true level of motivation. As outlined for H1a, employees' motivation is driven by the level of (positive or negative) reciprocity that stems from the employer's decision to allow or forbid telework. Thus, monitored office workers become more aware of their relatively low level of motivation (vis-à-vis teleworkers). Compared to less self-aware office workers (i.e., not monitored employees), monitored, self-aware office workers respond via a further disproportional reduction of motivation and effort. In other words, the effect of reciprocity on motivation is magnified under monitoring. This leads to our prediction that when monitoring is present rather than absent,

[32] Notably, while we argue that monitoring reduces autonomy, we have no theory to predict whether this effect is more pronounced for office vis-à-vis teleworkers. On the one hand, the effect might be stronger for teleworkers as they might perceive monitoring as an intrusion into their privacy (which is not the case in the office). On the other hand, the effect might be stronger for office workers as they work already in a more controlled environment (which is not the case for teleworkers) and surveillance might be perceived as particularly excessive in the office.

effort decreases more for office workers compared to teleworkers. This is formally stated in H2a.

H2a: When monitoring is present rather than absent, effort decreases more for office workers compared to teleworkers.

We refrain from predicting a main effect for monitoring for an important reason. While monitoring damages intrinsic motivation via reduced autonomy, it also makes teleworkers more self-aware of their reciprocity-driven high motivation (as predicted by H1a) which is likely to offset the negative effect. This is not the case for office workers resulting in the interaction effect predicted by H2a.

Our last hypothesis (H2b) focuses on the moderating effect of monitoring on misreporting. Contrary to effort, misreporting does not depend on motivation, but is rather a conscious decision. As we have no theory to predict that the crowding-out effect affects misreporting, we rely only on the self-awareness effect to derive our prediction.

Similar to H2a, we predict that monitoring increases self-awareness having different effects on office workers compared to teleworkers. However, as effort and honesty (or misreporting) are very different constructs with other drivers and antecedents, the mechanism that is caused by increased self-awareness differs. Above we argue that there is no salient and uniform standard that employees can compare their effort to and thus their current (low or high) level of intrinsic motivation is further intensified. A missing standard implies that employees exerting little effort will experience no dissonance between their current behavior and a standard that would redirect them towards a certain level of effort.

This is different for misreporting. Social norms generally expect employees to report honestly. As individuals usually share a preference for honesty (Evans et al. 2001), self-awareness enhances the salience of internal standards for honest behavior. While individuals may still lie to some extent to benefit from dishonest actions, self-awareness encourages them to critically assess whether such behavior aligns with their moral standards (Mazar et al. 2008). This self-focused attention leads individuals to critically evaluate the ethical implications of deviating from a commonly shared honesty standard, as deviating from their moral standards is likely to evoke psychological discomfort (Wicklund 1975). Consequently, self-aware individuals are generally less likely to misreport. This is also in line with prior research showing that self-awareness activates a moral evaluation process that discourages dishonest behavior (Cappelen et al. 2013; Fischbacher and Föllmi-Heusi 2013). Examples are honesty oaths or signing a declaration form that activate self-awareness and mitigate lying (Beck et al. 2018; Shu et al. 2012).

Building on the premise that self-aware employees are more likely to report honestly, we argue that the honesty-enhancing effect of monitoring is stronger for office workers compared to teleworkers.[33] As proposed by H1b, teleworkers engage in less misreporting compared to office workers in the absence of monitoring. Thus, office workers are further away from the standard of honest behavior. Consequently, when monitoring is present, the adjustment toward greater honesty is more pronounced for office workers compared to teleworkers. Based on these arguments, we formally propose our final hypothesis:

H2b: When monitoring is present rather than absent, misreporting decreases more for office workers compared to teleworkers.

Again, we refrain from predicting a main effect for monitoring. As just explained, teleworkers are expected to report more honestly than office workers (as argued by H1b). Thus, becoming more self-aware of the (natural) standard of honesty has little effect on teleworkers.

[33] We argue that the honesty-enhancing effect of self-awareness is stronger for office workers, but we still predict that teleworkers, overall, misreport less which is consistent with H1b. First, both office workers and teleworkers become more self-aware under monitoring, and there is no theory to suggest this effect is stronger for one group. However, the adjustment resulting from increased self-awareness is more pronounced for office workers, as they tend to engage in greater misreporting initially. Second, while self-awareness reduces the negative effect of misreporting, it is unlikely to produce a positive effect that increases honesty beyond the baseline observed in the absence of monitoring. Given teleworkers' higher baseline level of honesty, the increased self-awareness of office workers under monitoring will not surpass this level.

III. Research Design

Experimental Design and Procedure

To test our hypotheses, we conducted a computer-based experiment using a 2×2 between-subjects design. We manipulated monitoring (*present* vs. *absent*) and location (*office* vs. *telework*). The experiment was programmed and conducted using the SoPHIE software package (Hendriks 2012). The experimental procedure is depicted in Figure 1.

The experiment consisted of two parts, a pre-study and the main experiment. Ten participants were recruited for the pre-study to assume the role of employers (hereafter employers). Their task was to determine the working conditions for the participants in the main experiment (hereafter employees). Employers learned that some, but not all, employers were matched with employees in the main experiment who would work for them. During the pre-study, the employers answered a brief questionnaire where they made two decisions. First, employers decided whether "their" employees would have the possibility to telework or must work in the office (i.e., the university lab). Second, they determined whether their employees were monitored.[34] Employers were informed that their decisions could impact employee behavior, and that their compensation depended on employee behavior in the main task. On average, participants completed the questionnaire in 10 minutes and received a compensation of €2. Based on their choices on the working conditions, seven employers were allowed to participate in the main experiment.[35]

In the main experiment, 88 participants assumed the role of employees.[36] Each

[34] If employers opted for monitoring, they were asked whether they would personally monitor their assigned employees during the main experiment for an additional compensation of €15 or to have a third person doing this for them.

[35] Employers were informed that there is a possibility that they might not be selected to take part in the main experiment. When employers made the same choices, we randomly selected participants for the main experiment. To prevent gender effects, we ensured that a female and a male employer were chosen. However, due to employers' choices, only one employer could be assigned to an unmonitored treatment group, resulting in seven instead of eight employers in total. Since this treatment is unmonitored, gender effects (between the employer and the employees) could not materialize.

[36] In total, 91 business students participated in the main experiment. However, one student was excluded after trying to manipulate the source code of the experimental software (instead of focusing on the experiment),

employee was randomly assigned to one employer. In the first step, participants read the instructions and learned that their working conditions (location and monitoring) had been determined by another participant, i.e., the assigned employer, during the pre-study. Employees were also informed about the tasks and their compensation. After successfully responding to a quiz that ensured that participants' had correctly understood the instructions, they proceeded to the main tasks, i.e., an effort and a reporting task described below. To prevent order effects, an A/B split was implemented, randomly assigning participants to either begin with the effort or the reporting task.[37] Finally, participants completed a post-experimental questionnaire (PEQ) and were informed about their compensation. On average, the main experiment took 60 minutes and participants received a total compensation of €15.36. Compensation consists of a show-up fee of €3, a fixed compensation of €7 for the effort task, another fixed payment of €3 for the reporting task, and a variable payment based on decisions during the reporting task. Employees' behavior during the main task translated into an average compensation of €66.03 for their employers.[38]

[Place Figure 1 here]

Participants

The participants (i.e., employees) of the main experiment were 88 business students from a large Western European university. Their average age was 25.05 years, 38 (43.2%) were female and 61.4% had a bachelor's degree. There are no significant differences across conditions for age (p = 0.53, Kruskal-Wallis test), gender (p = 0.42, Chi-square test), educational degree (p = 0.35, Chi-square test), device used (e.g., touchpad or computer mouse) (p = 0.30, Chi-square test) and distance to the university (p = 0.46, Kruskal-Wallis test).[39]

and two other were excluded as they left the online conferencing platform which ensured monitoring.

37 There are no significant differences between group A and B (p > 0.1) in terms of our dependent variables effort and misreporting.

38 This compensation excludes the €15 show-up fee for monitoring employees.

39 All p-values are two-tailed.

Hence, randomization was successful. However, similar to Brüggen et al. (2024) who find that 70% of their participants have a preference to telework, 75% of our participants prefer working from home. Participants' preference for location showed significant differences between office workers and teleworkers.[40] Consequently, we control for location preference in our analyses.

Workplace and Monitoring Manipulation

We randomly assigned participants to the four treatment conditions of the main experiment. There was at least one employer assigned to each condition based on the employer's choices for the working conditions. Thus, employees knew that the employer they would be working for had determined their working conditions (location and monitoring).

We manipulated the working location at two levels: *office* vs. *telework*. Participants were randomly assigned to one condition and informed four days before the experiment where the experiment would take place. Participants in the *office* condition were instructed to come to the university lab at a specified time. Participants assigned to the *telework* condition participated from home and were also instructed to conduct the experiment at a specified time. All participants received a link and logged on to the identical experimental website. Hence, apart from the physical working location we kept the treatments as consistent as possible, including the time participants conducted the experiment. Following prior research, the laboratory appears to be an adequate office operationalization. The campus can be understood as the students' natural working environment where they regularly attend lectures. Participating from home—away from this central workplace—aligns with the definition of telecommuting and can be considered telework.

The second factor is monitoring (*present* vs. *absent*). Participants in the monitoring-present condition had to log on to a video conferencing system while working on the two main

[40] We ask participants in the PEQ on a 7-point Likert scale to what extent they agree with the statement: "I prefer to complete work assignments at home rather than at the university." (7-point Likert scale, reversed). Teleworkers agree significantly more with the statement compared to office workers (5.60 vs. 4.73; $t = -2.63$; $p = 0.01$).

tasks. Instructions on how to sign in were provided and participants were required to switch on their camera and microphone. If the assigned employer was able to see and hear the participants, they received a password via chat that allowed them to proceed to the main tasks. Our operationalization of monitoring was intended to capture the feature of surveillance tools used in practice that monitor physical presence. Each participant entered an individual conferencing room with the employer present and visible but not the peers. After finishing both tasks and before answering the PEQ, employees left the meeting. Employers documented whether employees remained in the meeting throughout.[41] Participants in the monitoring-absent condition did not join a conferencing room.

Effort Task

We used the slider task to measure effort (Araujo et al. 2016; Brüggen et al. 2024; Chan 2018; Gill and Prowse 2012). During this task, participants had to solve as many sliders as possible for 30 minutes. Contrary to previous research, we set the time for the task rather long for several reasons. First, it is often assumed that teleworkers do not spend their entire working time on the task assigned due to high opportunity costs when working from home. We argue that this effect will only materialize when the task takes more than just two or four minutes (Araujo et al. 2016; Brüggen et al. 2024; Chan 2018; Gill and Prowse 2019).[42] Second, while participants might find it engaging to work on the task for a short period, it requires much more intrinsic motivation to work on it for 30 minutes. Based on our theory, intrinsic motivation is a crucial determinant, thus we intentionally designed the task to ensure that intrinsic motivation varies.

[41] From the three participants excluded in total, two were excluded due to non-compliance with these instructions, i.e., they both left the meeting—which ensured monitoring—after receiving the password to continue. Including the two (or three) subjects yields inferentially identical results.

[42] This is in line with Brüggen et al. (2024) who report no significant differences in effort among teleworkers and office workers on their 4-minute slider task. Similarly, we find no significant differences in effort when monitoring is absent during the first four minutes of the task ($F = 0.96$, $p = 0.48$, two-tailed). Notably, significant effects only emerge after doubling the task duration to 8 minutes and onwards ($F = 2.20$, $p = 0.05$, two-tailed).

Given the fact that office workers can also engage in private activities at work (e.g., coffee breaks with colleagues), even though their opportunities are limited, two design choices were made. First, participants in all conditions could take a break from the slider task for up to ten times and play tic-tac-toe for 30 seconds against the computer (e.g., Dutcher 2012). During the tic-tac-toe sessions, participants could not solve any sliders. Second, participants in all treatments were informed that they could use their mobile phones during the experiment.

During the task, 45 sliders arranged across three columns were presented on screen. All sliders were initially positioned at 0. Participants were asked to move as many sliders as possible to the target position of 50. Every 2 minutes, the screen refreshed, and 45 new sliders appeared.[43] To prevent participants from developing strategies based on the slider arrangement on the screen, the sliders within each column were staggered on every page (Gill and Prowse 2019). Additionally, participants could not use the arrow keys on the keyboard to move the sliders. To avoid effects of using different devices, participants had to indicate whether they have and use a computer mouse or a touchpad for their daily work when signing up for the experiment. We used this information to make sure that the percentage of touchpad/mouse users is the same in the office and telework treatments.[44] We also ensured uniformity in the display of sliders by calibrating the pixel frame to various screen sizes.[45] While employee-participants received a fixed compensation of €7 for solving as many sliders as possible, the assigned employers' compensation depended on the number of sliders solved. In detail, a random mechanism determined the remuneration per slider which varied between 10 and 15 lira.[46] This feature—communicated to employers and employees—made sure that there was no perfect

[43] A pretest revealed that it is impossible to solve 45 sliders within 2 minutes. This design choice ensured that an unlimited number of sliders was available.

[44] Based on the responses, we equipped about 44% of workplace participants with touchpads. In the telework treatment, about 47% of participants ended up using a touchpad. There are no significant differences between both treatment groups in terms of the device used for the slider task.

[45] This is particularly relevant as participants in the telework condition use their private computer which can differ in screen size. None of the participants used a tablet or smartphone.

[46] In the experiment, compensation was denominated in the experimental currency lira. At the end of the experiment, lira was converted into euros with 1,000 lira equaling one euro.

monitoring as employers could not use compensation information from the slider task to calculate employee performance. Our main dependent variable used to test the effort hypotheses (H1a and H2a) is the number of sliders moved.[47]

Reporting Task

We measure misreporting using the reporting task by Evans et al. (2001). The task builds on the budgeting task by Antle and Eppen (1985) and has become widely used for assessing honesty (Brüggen et al. 2024; Farrell et al. 2021; Hannan et al. 2006). A typical reporting activity in the real world is budgetary reporting (Brown et al. 2009; Webb 2002), which requires employees to reveal private information to support planning and resource allocation (Antle and Fellingham 1997; Sprinkle 2003). Similarly, many employees report their working hours or time spent on specific tasks or projects (Agoglia et al. 2015; Reid 2015). Such situations often entail opportunities and incentives for dishonesty (Church et al. 2012; Hannan et al. 2006; Rankin et al. 2008).

In the reporting task, participants submitted cost reports for four projects. To avoid the end-of-game effect, employees did not know the number of projects ex-ante (Farrell et al. 2021). Each project generated revenues of 2,000 lira (experimental currency) (Brüggen et al. 2024). The actual project costs were randomly drawn but equal for all participants (Church et al. 2012), i.e., 1,000 lira, 200 lira, 800 lira, and 1,700 lira.

Employees could report any value between the actual cost and 2,000 lira (i.e., the project revenue). They knew that the reported value would be accepted, and that the employer would never learn the actual costs. The difference between the reported cost and the actual cost increased employees' payoff. Hence, employees had an incentive to misreport by overstating the project costs. At the same time, this overstatement reduced the employer's compensation.

[47] This proxy for effort satisfies the three criteria for effort established by Baiman (1982): the individual must have control over the proxy which must be correlated with performance. Finally, the proxy must be costly. These criteria are fulfilled. Participants have control over moving the slider, moving the sliders to the target position is associated with performance and the opportunity costs (e.g., using the mobile phone or playing tic-tac-toe) make effort costly.

Employers received the difference between the project revenue and the reported cost.[48]

In addition to a fixed pay of €3 for the reporting task, employees earned on average a compensation of €2.36 from misreporting. Consistent with prior literature (Brüggen et al. 2024; Church et al. 2012; Evans et al. 2001), we measure misreporting (our second dependent variable) by dividing the amount participants misreported by the maximum possible amount for misreporting:

$$Misreporting = \frac{\sum_{k=1}^{4} reported\ cost\ k\ - actual\ cost\ k}{\sum_{k=1}^{4} 2{,}000\ - actual\ cost\ k}$$

Misreporting thus ranges from 0 (indicating complete honesty) to 1 (indicating complete dishonesty).

IV. RESULTS

Descriptive Results and Hypotheses Tests

Table 1 depicts the mean and standard deviation for our two main dependent variables, i.e., *effort* and *misreporting*. Figure 2 depicts a graphical representation. Effort is measured by the total number of sliders moved and misreporting by the extent to which participants misreport their cost information during the reporting task.

H1a predicts that effort is higher for teleworkers compared to office workers. In line with this prediction, the descriptive statistics reveal that the number of sliders moved, i.e. effort, is greater in the telework condition (333.98) than in the office condition (246.85). To formally test H1a, we use analysis of covariance (ANCOVA). We control for devices used during the slider task and location preference for two reasons: First, the time to adjust a slider in the effort task depends on the device used (i.e., mouse or touchpad). To avoid an interference with our measurement, we include a dummy variable for device. Second, participants in the telework

[48] To help employees with the calculation, a pop-up box appeared before submitting a cost report. The actual project cost, the reported cost, and the resulting compensation for the employee and the employer was displayed (Farrell et al. 2021).

condition exhibit a stronger preference for telework, thus randomization was not completely successful.[49] Thus, we control for location preference.[50] Based on the descriptive results and the ANCOVA, we conclude that effort is higher for teleworkers compared to office workers. The results in Table 2, Panel A confirm that effort is significantly greater for teleworkers compared to office workers ($F = 17.12$; $p < 0.01$).[51] Thus, H1a is supported.

H1b predicts that misreporting is lower for teleworkers compared to office workers. The descriptive results in Table 1, Panel B show that misreporting is lower for teleworkers (0.51) compared to office workers (0.59), i.e., while teleworkers ask for 51% of the available budget (project revenue – actual costs), office workers demand 59%. We formally test H1b using ANCOVA. Again, we include location preference as a control variable. In addition to our argument from above, i.e., failed randomization, another reason for including this control exists: Brüggen et al. (2024) report that teleworkers with a preference for telework are less honest. Thus, it is important to control for this effect. Contrary to our test for H1a, we do not control for device, as device matters for the slider task but not for the reporting task. The corresponding ANCOVA results are displayed in Table 3, Panel A and confirm that teleworkers misreport significantly less ($F = 4.02$; $p = 0.05$). Hence, H1b is supported.

[Place Table 1 here]

Next, we examine the moderating effect of monitoring. H2a predicts that when monitoring is present, the decrease in effort is greater for office workers compared to teleworkers. The descriptive results in Table 1, Panel A show that the number of sliders moved, i.e., our proxy for effort, decreases by 37.77 from 266.20 (monitoring absent) to 228.43 (monitoring present) for office workers. For teleworkers, the decrease is only 9.75 (338.96

49 An analysis of variance (ANOVA) using location preference as the dependent variable and treatment as the independent variable shows significant differences between the four treatment groups ($p = 0.09$, untabulated). Hence, randomization was not completely successful.

50 The results for our hypotheses are inferentially identical if we do not control for location preference. Yet, the p-value for H1b is slightly above conventional levels of significance ($F = 1.96$; $p = 0.17$, two-tailed), or only significant when one-tailed p-values are used.

51 All p-values are reported two-tailed unless stated otherwise.

when monitoring is absent to 329.21 when monitoring is present).

To formally test the pattern predicted by H2a, we use the nonparametric Jonckheere-Terpstra test and planned contrast analysis. The conventional ANCOVA analysis in Table 2 shows an insignificant interaction effect *Location* × *Misreporting* ($p = 0.46$). Yet, while ANOVA is suitable for identifying disordinal interactions, it is less effective in detecting (semi)ordinal interactions (Buckless and Ravenscroft 1990). Since our hypothesis predicts an ordinal interaction, consistent with the graphical representation in Figure 2, we first apply the conservative nonparametric Jonckheere-Terpstra test. This test does not require specifying contrast weights. The results support H2a ($Z = 4.08$; $p < 0.01$; untabulated). Next, we follow Buckless and Ravenscroft (1990) and use contrast analysis with specified weights. According to our prediction, we use contrast weights of -1 for *Office/Monitoring absent* and -7 for *Office/Monitoring present* showing a greater decrease in effort for office workers compared to teleworkers for whom we use contrast weights of +5 for *Telework/Monitoring absent* and +3 for *Telework/Monitoring present,* as self-awareness is argued to partially offset the autonomy effect for teleworkers. Table 2, Panel B reports the results, which again support H2a ($F = 19.27$; $p < 0.01$).[52] Hence, we conclude that when monitoring is present, the effort decrease is more pronounced for office workers than for teleworkers, supporting H2a.

[Place Table 2 and Figure 2 here]

Our last hypothesis H2b predicts that when monitoring is present, the decrease in misreporting is greater for office workers compared to teleworkers. The descriptive results in Table 1, Panel B are in line with this prediction: While misreporting decreases by 25% in the office condition (monitoring absent: 0.68 vs. monitoring present: 0.51), it declines by only 8%

52 We follow the three-step approach suggested by Guggenmos et al. (2018) for planned contrasts: 1) our contrast weights visually fit the descriptive data (Figure 2, Panel A), 2) our data significantly match the planned contrast weights ($F = 19.27$; $p < 0.01$) and 3) the q^2 is 0.04, hence the contrast explains 96% of the between-cell variance. Our results for H2a are robust ($F = 19.41$; $p < 0.01$) to using alternative contrast weights of -1 for *Office/Monitoring absent*, -5 for *Office/Monitoring present,* +3 for *Telework/Monitoring absent,* and +3 for *Telework/Monitoring present*.

in the telework condition (monitoring absent: 0.53 vs. monitoring present: 0.49). To formally test the pattern predicted by H2b, we again rely on the nonparametric Jonckheere-Terpstra test and planned contrast analysis for the reasons stated above. Without specifying contrast weights, the results of the Jonckheere-Terpstra test support H2b ($Z = 2.14$; $p < 0.03$; untabulated). For the contrast analysis we use contrast weights of +4 for *Office/Monitoring absent* and -1 *Office/Monitoring present* implying a greater decrease in misreporting compared to teleworkers for which we use contrast weights of -1 for *Telework/Monitoring absent* and -2 for *Telework/Monitoring present* as teleworkers have little reason to adapt misreporting in response to increased self-awareness. As shown in Table 3, Panel B, H2b is supported ($F = 7.58$; $p < 0.01$).[53]

It follows from H2a and H2b that monitoring has a stronger effect on office workers compared to teleworkers. This means that effort and misreporting decrease more when monitoring is present for office workers compared to teleworkers. However, it is noteworthy that from the employer's perspective monitoring has a negative effect on office workers' effort (less effort) but a positive effect on misreporting (less misreporting).

[Place Table 3 here]

Additional Analysis

This subsection presents additional analyses and explores responses from the PEQ to further evaluate our theoretical framework. More precisely, we validate the core assumptions underlying our experimental design and provide process evidence in support of our reciprocity-based explanation for H1 and H2.

[53] We follow again the three-step approach suggested by Guggenmos et al. (2018) for planned contrasts: 1) our contrast weights visually fit the descriptive data (Figure 2, Panel B), 2) our data significantly match the planned contrast weights ($F = 7.99$; $p < 0.01$) and 3) the q^2 is 0.01, hence the contrast explains 99% of the between-cell variance. Our results for H2b are robust ($F = 7.58$; $p < 0.01$) to alternative weights of +3 for *Office/Monitoring absent*, -1 for *Office/Monitoring present,* -1 for *Telework/Monitoring absent,* and -1 for *Telework/Monitoring present.*

Validation of Experimental Design

To validate our experimental design, we first assess whether our monitoring manipulation was successful. Next, we test whether our core assumption for reciprocity, i.e., that participants prefer telework, is met.

In the PEQ, we asked participants on a 7-point Likert scale to what extent they agree with the statement: "I felt monitored during the experiment." (7-point Likert scale, reversed). Participants in the monitoring-present condition (mean 4.8) felt significantly more monitored ($t = 9.74$; $p < 0.01$) than participants in the monitoring-absent condition (mean 1.84). Thus, we conclude that our monitoring manipulation was successful.

To assess participants' preference for location, we asked to what extent they agree with the statement: "I prefer to complete work assignments at home rather than at the university." (7-point Likert scale, reversed). Teleworkers agree significantly more with the statement compared to office workers (5.60 vs. 4.73; $t = -2.63$; $p = 0.01$). Notably, participants in both conditions express a clear preference for telework, as the mean scores in each group are significantly above the neutral midpoint of 4 (telework condition: $t = 8.23$; $p < 0.01$; office condition: $t = 2.68$; $p = 0.01$), indicating a pronounced preference for telework.

Process Evidence for Effort

Next, we conduct a path analysis to gain further insights into our theory. The path model for effort is depicted in Figure 3, Panel A and reflects the test of our theory for H1a and H2a. H1a predicts that teleworkers exert more effort than office workers. In our theory development we explain that this direct effect between location and effort (path c') is mediated by reciprocity and motivation. We argue that location affects reciprocity (path a_2) driving motivation (path d_2). Motivation determines effort (path b).

H2a predicts that under monitoring effort decreases more for office workers compared to teleworkers. The model in Figure 3 shows that monitoring is expected to affect motivation via two channels: First, by reducing autonomy (path a_1), which in turn reduces motivation (path

d_1) and second, by moderating the mediating effect of reciprocity on motivation through increased self-awareness (path ii). We construct the three mediators—autonomy, motivation and (effort) reciprocity—in the model by performing principal component analysis on related items from the PEQ (Table 4).[54]

[Place Table 4 here]

We use the PROCESS macro in SPSS to test our path model including the moderated-mediation (Hayes 2018). We test the significance of each mediation effect by estimating bias-corrected bootstrap confidence intervals with 10,000 resamples (Hayes 2018).

As predicted for H1a, we find that location (with telework coded as 1 and 0 otherwise) affects reciprocity. More precisely, we find that teleworkers (office workers) reciprocate (retaliate) their employer (ß = 0.88; t = 4.55; $p < 0.01$) which increases (decreases) motivation (ß = 0.28; t = 2.13; p = 0.04). Higher levels of motivation increase effort (ß = 22.42; t = 2.04; p = 0.04). The indirect effect of location on effort (through reciprocity and motivation) is significant, irrespective of whether monitoring is absent (the 90 percent CI of 0.28 to 17.15 does not include 0) or present (the 90 percent CI of 2.17 to 33.49 does not include 0). For H2a, we rely on self-awareness and predict that this indirect effect is moderated by monitoring. In fact, we find support for the predicted, moderated mediation (the 90 percent CI of 1.28 to 24.43 does not include 0). The effect of location on effort via the mediators' reciprocity and motivation is 13.71 when monitoring is present and 5.59 when it is absent.

As predicted, we find that monitoring reduces autonomy (ß = -1.30; t = -2.85; $p < 0.01$). We also find that teleworkers perceive higher autonomy compared to office workers (ß = 0.87;

[54] The reciprocity factor for effort has an eigenvalue of 1.87 and a Cronbach's alpha of 0.68, which is only slightly below the conventional level of 0.7 (for a discussion of allowance for broader ranges of Cronbach's alpha see DeZoort and Salterio (2001); Koch and Salterio (2017); Tan and Kao (1999)). The Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy is 0.58, meeting the minimum threshold for factor analysis (Kaiser 1970; Stewart 1981). The scale ranges from –2.85 (retaliation) to 1.23 (reciprocity), with higher values reflecting greater willingness to reciprocate. The motivation factor for effort yields an eigenvalue of 1.65, a Cronbach's alpha of 0.78, and a KMO of 0.50, indicating high internal consistency. Autonomy was measured using a single item and is therefore not included in the reliability assessment.

t = 1.94; p = 0.06). The effect of autonomy on motivation is negative, as expected, but not significant (ß = -0.05; t = -0.79; p = 0.43).[55] Hence, it is only reciprocity that determines motivation. Interestingly, teleworkers' autonomy is particularly damaged under monitoring (ß = -1.30; t = -2.85; p < 0.01). This may be because teleworkers perceive surveillance as a violation of their privacy. Overall, the path model generally supports our theory for H1a and H2a.

Process Evidence for Misreporting

We use a similar path model to validate our theory for misreporting (H1b and H2b). For this model, we consider only (misreporting) reciprocity as a mediator, as (dis)honest reporting constitutes a conscious decision and is not affected by intrinsic motivation.[56] The model is shown in Figure 3, Panel B.

[Place Figure 3 here]

We again find that teleworkers are significantly more likely to positively reciprocate compared to office workers (ß = 0.74; t = 3.68; p < 0.01). Further, we find a significant negative effect of reciprocity on misreporting (ß = -0.09; t = −1.94; p = 0.06). Thus, the willingness to reciprocate (retaliate) decreases (increases) individuals' tendency to misreport. Location has no direct effect on misreporting (ß = -0.11; t = −1.18; p = 0.24).[57] Instead, the willingness to reciprocate fully mediates the relation between location and misreporting. Specifically, telework, compared to office work, results in less misreporting in the absence of monitoring, with a significant indirect effect of -0.07 (the 90 percent CI of -0.13 to -0.01 does not include 0). This pattern also holds true under monitoring, with an indirect effect of -0.05. However, the

[55] The indirect effect of location through autonomy and motivation on effort is also insignificant. A potential reason why we fail to find a significant effect of autonomy on intrinsic motivation is that our effort task is rather low in intrinsic motivation. In this vein, Ryan and Deci (2000) argue that the effects predicted by social determination theory (and related sub-theories) depend on task characteristics.

[56] The reciprocity factor for misreporting has an eigenvalue of 1.92 and a Cronbach's alpha of 0.70. The KMO is 0.60. The resulting scale ranges from –3.01 (retaliation) to 1.31 (reciprocity), with higher values indicating greater willingness to reciprocate.

[57] When applying an ANCOVA and controlling for participants' preference of location, we find a significant effect of location on misreporting as predicted by H1b (F = 4.02; p = 0.05).

90% confidence interval (–0.14 to 0.00) includes zero, placing the effect just above the conventional threshold for significance.[58] This pattern is directionally consistent with our theory for H1b.

For H2b we argue that the effect of location on misreporting via the mediator reciprocity depends on monitoring. More precisely, we argue that the effect of reciprocity (particularly of negative reciprocity) is weakened when monitoring increases self-awareness. We find that the decrease in misreporting for teleworkers compared to office workers is lower when monitoring is present (-0.05) rather than absent (-0.07). Put differently, under monitoring, the location effect on misreporting is reduced. However, the moderated mediation effect is insignificant (the 90 percent CI of -0.08 to 0.09 includes 0). This suggests that there might be other factors that explain the decrease in misreporting under monitoring. Since the direct effect of monitoring on misreporting is negative and significant ($ß = -0.18$; $t = -2.09$; $p = 0.04$), being monitored is likely to increase self-awareness that results in an alignment with individuals' internal moral standards. Surprisingly, this even holds when monitoring is ineffective. This aligns with results by Bateson et al. (2006), who find that the mere display of a pair of watching eyes on a piece of paper increases honest behavior.

Hence, while we provide process evidence that reciprocity mediates the relation between location and misreporting, future research is needed to explore the effects of self- awareness on teleworkers and office workers in the presence of monitoring. Moreover, our findings show that monitoring increases office workers' tendency to retaliate by reducing effort, while simultaneously decreasing their tendency to retaliate through dishonest behavior. It is

[58] When using a confidence interval of 88%, the indirect effect becomes significant (CI of -0.14 to -0.01 does not include 0). Similarly, when refining the reciprocity factor by removing the item capturing participants' resentment toward their working location, the indirect effect of -0.04 is again significant (the 90 percent CI of -0.10 to -0.01 does not include 0). This adjusted two-item measure has a Cronbach's alpha of 0.54. While this indicates modest internal consistency, it aligns with prior accounting research using short or composite measures that reflect related but distinct behavioral dimension (Dearman and Shields 2001; Tan and Kao 1999). Moreover, Cronbach's alpha assumes unidimensionality and can underestimate reliability for constructs that encompass opposing motivations—such as prosocial reciprocity versus disengagement.

important to distinguish between retaliation via reduced effort—driven by low motivation—and intentional dishonesty, such as actively deceiving or stealing from an employer.

V. Conclusion

Today's working environment is shaped by two megatrends: telecommuting and employee surveillance. While both are widespread in practice, research on how telecommuting affects employee behavior and how it interacts with monitoring is still scarce. To answer these research questions via an experiment, we first examine effort and misreporting of teleworkers vis-à-vis office workers. Second, we investigate the moderating role of monitoring. To increase external validity, we focus on a setting where employees are aware that it is the employer's deliberate decision to allow or deny telework and to implement monitoring or not. Employees can directly reciprocate or retaliate against the employer for these decisions.

Building on reciprocity, we predict and find that teleworkers exhibit greater effort and misreport less than office workers. Under monitoring, effort and misreporting decrease more for office workers compared to teleworkers. For effort, we argue that monitoring increases self-awareness and thereby intensifies the motivational effects of reciprocity. For misreporting, increased self-awareness makes particularly office workers aware that misreporting contradicts well-established preferences for honesty. Thus, the effect of location on misreporting is weakened under monitoring.

Path models confirm that the different levels of effort and misreporting of teleworkers vis- à-vis office workers result from differences in reciprocity. The models also lend further support for our argument that monitoring moderates the mediating effect of reciprocity on motivation and effort.

Our findings may help organizations design the working environment. First, we show that teleworkers reciprocate employer's trust for allowing remote work by more effort and less misreporting. On the other hand, office workers retaliate employers for being denied the option to telework. They exert less effort and misreport more. These findings are important as employers are concerned about opportunistic behavior of teleworkers (e.g., Tesla, Porter 2022), and an increasing number of firms responded with return-to-office policies (Business Insider 2024a). Our results suggest that firms should be aware of the high costs associated with this decision.

Second, our study implies that firms must carefully distinguish whether office workers' performance heavily depends on intrinsic motivation (e.g., effort), where monitoring has negative effects, or is less dependent on intrinsic motivation (e.g., misreporting), where monitoring may yield positive effects. Third, our research setting, which incorporates reciprocity and an ineffective monitoring system, closely mirrors real-world conditions, making our results more generalizable for practical applications.

Future research could further explore this field of research. While our study strictly differentiates between telework and office work, firms also adopt hybrid models in practice. Future research might examine such hybrid work arrangements or investigate settings where teleworkers are forced to return to the office. Further, we investigate the effects of telecommuting and monitoring via an effort and a reporting task. While both tasks are common for many jobs, it would be interesting to explore further performance dimensions. Examples are creativity tasks, collaborative group tasks, or problem-solving tasks that emphasize critical thinking. Lastly, we purposely implement imperfect monitoring as increasingly observed in practice (e.g., Ball 2021). However, future research could implement different types or even a combination of several surveillance systems. Moreover, we show that the perception of fairness of an employer's decision plays a crucial role in determining an employee's level of reciprocity. Hence, small design choices might influence this perception, such as providing an explanation of why monitoring might be beneficial for the employee (i.e., to prevent overwork).

References

Agoglia, C.P., R. C. Hatfield, and T. A. Lambert 2015. Audit team time reporting: An agency theory perspective. *Accounting, Organizations and Society* 44: 1–14. https://doi.org/10.1016/j.aos.2015.03.005.

Akerlof, G.A. 1982. Labor Contracts as Partial Gift Exchange. *The Quarterly Journal of Economics* 97 (4): 543–569. https://doi.org/10.2307/1885099.

Aksoy, C.G., J. M. Barrero, N. Bloom, S. J. Davis, M. Dolls, and P. Zarate 2023. Working from Home Around the Globe: 2023 Report. *EconPol Policy Brief* 53.

Alge, B.J., Hansen, D. (Eds.) 2013. Workplace Monitoring and Surveillance Research since “1984”: A Review and Agenda, 1st ed. Routledge, New York.

Allen, T.D., T. D. Golden, and K. M. Shockley 2015. How effective is telecommuting? Assessing the status of our scientific findings. *Psychological science in the public interest* 16 (2): 40–68. https://doi.org/10.1177/1529100615593273.

Amazon 2024. Message from CEO Andy Jassy: Strengthening our culture and teams. https://www.aboutamazon.com/news/company-news/ceo-andy-jassy-latest-update-on-amazon-return-to-office-manager-team-ratio.

American Psychological Association 2023. 2023 Work in America Survey. https://www.apa.org/pubs/reports/work-in-america/2023-work-america-ai-monitoring.

Antle, R., and G. D. Eppen 1985. Capital Rationing and Organizational Slack in Capital Budgeting. *Management Science* 31 (2): 163–174. https://doi.org/10.1287/mnsc.31.2.163.

Antle, R., and J. Fellingham 1997. Models of Capital Investments with Private Information and Incentives: a Selective Review. *Journal of Business Finance & Accounting* 24 (7-8): 887–908. https://doi.org/10.1111/1468-5957.00140.

Araujo, F.A., E. Carbone, L. Conell-Price, M. W. Dunietz, A. Jaroszewicz, R. Landsman, D.

Lame, L. Vesterlund, S. W. Wang, and A. J. Wilson 2016. The slider task: an example of restricted inference on incentive effects. *Journal of the Economic Science Association* 2: 1–12. https://doi.org/10.1007/s40881-016-0025-7.

Baiman, S. 1982. Agency research in management accounting: a survey. *Journal of Accounting Literature* 1 (1): 154–210. https://doi.org/10.1016/0361-3682(90)90023-N.

Ball, K. 2021. Electronic Monitoring and Surveillance in the Workplace. *European Commission Joint Research Centre*. htpps://doi.org/10.2760/5137.

Bateson, M., D. Nettle, and G. Roberts 2006. Cues of being watched enhance cooperation in a real-world setting. *Biology Letters* 2 (3): 412–414. https://doi.org/10.1098/rsbl.2006.0509.

BBC 2020. Barclays scraps 'Big Brother' staff tracking system. https://www.bbc.com/news/business-51570401.

Beck, T., C. Bühren, B. Frank, and E. Khachatryan 2018. Can Honesty Oaths, Peer Interaction, or Monitoring Mitigate Lying? *Journal of Business Ethics* 163 (3): 467–484. https://doi.org/10.1007/s10551-018-4030-z.

Belanger, F. 1999. Workers' propensity to telecommute: An empirical study. *Information & Management* 35: 139–153. https://doi.org/10.1016/S0378-7206(98)00091-3.

Bloom, N., J. Liang, J. Roberts, and Z. J. Ying 2015. Does working from home work? Evidence from a chinese experiment. *The Quarterly Journal of Economics* 130 (1): 165. https://doi.org/10.3386/w18871.

Brown, E. 2020. Employee surveillance software demand increased as workers transitioned to home working. *ZDNet*. https://www.zdnet.com/article/employee-surveillance-software-demand-increased-as-workers-transitioned-to-home-working/.

Brown, J.L., J. H. Evans, and D. V. Moser 2009. Agency Theory and Participative Budgeting Experiments. *Journal of Management Accounting Research* 21 (1): 317–345. https://doi.org/10.2308/jmar.2009.21.1.317.

Brüggen, A., C. Feichter, and K. Haesebrouck 2024. Home Office: Evidence on Location and Selection Effects of Telecommuting. *Journal of Management Accounting Research* 36 (3): 95–113. https://doi.org/10.2308/JMAR-2023-053.

Buckless, F.A., and S. P. Ravenscroft 1990. Contrast Coding: A Refinement of ANOVA in Behavioral Analysis. *The Accounting Review* 65 (4): 933–945.

Business Insider 2024a. The list of major companies requiring employees to return to the office, from The Washington Post to Amazon. https://www.businessinsider.com/companies-requiring-return-to-office-rto-mandate.

Business Insider 2024b. Wegen Unmut über weniger Homeoffice: Deutsche Bank soll Kommentarfunktion im Intranet deaktiviert haben. https://www.businessinsider.de/wirtschaft/deutsche-bank-petition-gegen-die-einschraenkung-des-homeoffice/.

Camerer, C.F. 2011. Behavioral game theory: Experiments in strategic interaction. Princeton university press, Princeton.

Canon 2024. Workplace AI. https://asia.canon/en/business/workplace-ai/product.

Cappelen, A.W., E. Ø. Sørensen, and B. Tungodden 2013. When do we lie? *Journal of Economic Behavior & Organization* 93: 258–265. https://doi.org/10.1016/j.jebo.2013.03.037.

Carver, C.S. 2012. Self-Awareness, in Leary, M.R., Tangney, J.P. (Eds.), Handbook of Self and Identity. The Guilford Press, New York, 50-68.

Chan, E.W. 2018. Promotion, relative performance information, and the peter principle. *The Accounting Review* 93 (3): 83–103.

Chen, R. 2024. Amazon RTO: Employees Don't Want to Work in the Office Five Days a Week. *Blind Workplace Insights*. https://www.teamblind.com/blog/amazon-rto-return-to-office-five-days-week-reaction/.

Christ, M.H., S. A. Emett, S. L. Summers, and D. A. Wood 2012. The Effects of Preventive and Detective Controls on Employee Performance and Motivation. *Contemporary Accounting Review* 29 (2): 432–452. https://doi.org/10.1111/j.1911-3846.2011.01106.x.

Christ, M.H., K. L. Sedatole, K. L. Towry, and M. A. Thomas 2008. When Formal Controls Undermine Trust and Cooperation. *Strategic Finance* 89 (7): 38–44.

Church, B.K., R. L. Hannan, and X. J. Kuang 2012. Shared interest and honesty in budget reporting. *Accounting, Organizations and Society* 37 (3): 155–167. https://doi.org/10.1016/j.aos.2012.01.002.

Da Silva, A.D., D. Georgarakos, and M. Weißler 2023. How people want to work–preferences for remote work after the pandemic. *Economic Bulletin* 1.

Davidson, B.I. 2019. The Effects of Reciprocity and Self-Awareness on Honesty in Managerial Reporting: Social Value Orientation Matters. *Journal of Management Accounting Research* 31 (1): 85–103. https://doi.org/10.2308/jmar-52020.

Dearman, D.T., and M. D. Shields 2001. Cost Knowledge and Cost-Based Judgment Performance. *Journal of Management Accounting Research* 13 (1): 1–18.

Deci, E., and R. M. Ryan 2012. Self-determination theory. *Handbook of theories of social psychology* 1 (20): 416–436.

DeZoort, F.T., and S. E. Salterio 2001. The Effects of Corporate Governance Experience and Financial-Reporting and Audit Knowledge on Audit Committee Members' Judgments. *AUDITING: A Journal of Practice & Theory* 20 (2): 31–47.

Dickinson, D., and M.-C. Villeval 2008. Does monitoring decrease work effort? The complementarity between agency and crowding-out theories. *Games and Economic Behavior* 63 (1): 56–76. https://doi.org/10.1016/j.geb.2007.08.004.

Dutcher, E.G. 2012. The effects of telecommuting on productivity: An experimental examination. The role of dull and creative tasks. *Journal of Economic Behavior & Organization* 84 (1): 355–363. https://doi.org/10.1016/j.jebo.2012.04.009.

Enzle, M.E., and S. C. Anderson 1993. Surveillant Intentions and Intrinsic Motivation. *Journal of Personality and Social Psychology* 64 (2): 257–266. https://doi.org/10.1037/0022-3514.64.2.257.

European Digital Rights 2022. WFH – Watched from Home: Office 365 and workplace surveillance creep. https://edri.org/our-work/wfh-watched-from-home-office-365-and-workplace-surveillance-creep/.

Evans, J.H., R. L. Hannan, R. Krishnan, and D. V. Moser 2001. Honesty in Managerial Reporting. *The Accounting Review* 76 (4): 537–559. https://doi.org/10.2139/ssrn.269448.

Falk, A., and U. Fischenbacher 2006. A theory of reciprocity. *Games and Economic Behavior* 54 (2): 293–315. https://doi.org/10.1016/j.geb.2005.03.001.

Farrell, A.M., J. H. Grenier, and J. Leiby 2017. Scoundrels or Stars? Theory and Evidence on the Quality of Workers in Online Labor Markets. *The Accounting Review* 92 (1): 93–114. https://doi.org/10.2139/ssrn.2424718.

Farrell, A.M., S. A. Peffer, K. Rotaru, and A. K. D. Schulz 2021. The Effects of Forecast Precision and Time Pressure on Motives to Create Budget Slack. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3874286.

Fessler, N.J. 2003. Experimental Evidence on the Links among Monetary Incentives, Task Attractiveness, and Task Performance. *Journal of Management Accounting Research* 15 (1): 161–176. https://doi.org/10.2308/jmar.2003.15.1.161.

Fischbacher, U., and F. Föllmi-Heusi 2013. Lies in Disguise: An Experimental Study on Cheating. *Journal of the European Economic Association* 11 (3): 525–547. https://doi.org/10.1111/jeea.12014.

Fortune Business Insights 2024. Employee Surveillance and Monitoring Software Market. https://www.fortunebusinessinsights.com/employee-surveillance-and-monitoring-software-market-104796.

Frey, B.S., and R. Jegen 2001. Motivation Crowding Theory. *Journal of Economic Surveys* 15 (5): 589–611. https://doi.org/10.1111/1467-6419.00150.

Gagne, M., and E. L. Deci 2005. Self-determination theory and work Motivation. *Journal of Organizational Behavior* 26: 331–362. https://doi.org/10.1002/job.322.

Gajendran, R.S., and D. A. Harrison 2007. The good, the bad, and the unknown about telecommuting: meta-analysis of psychological mediators and individual consequences. *The Journal of applied psychology* 92 (6): 1524–1541. https://doi.org/10.1037/0021-9010.92.6.1524.

Gibbs, M., F. Mengel, and C. Siemroth 2021. Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals. *SSRN Electronic Journal*. https://dx.doi.org/10.2139/ssrn.3841567.

Gill, D., and V. Prowse 2012. A Structural Analysis of Disappointment Aversion in a Real Effort Competition. *American Economic Review* 102 (1): 469–503. https://doi.org/10.1257/aer.102.1.469.

Gill, D., and V. Prowse 2019. Measuring costly effort using the slider task. *Journal of Behavioral and Experimental Finance* 21: 1–9. https://doi.org/10.1016/j.jbef.2018.11.003.

Global Workplace Analytics 2021. Latest Work-at-Home/Telecommuting/Remote Work Statistics. https://globalworkplaceanalytics.com/telecommuting-statistics.

Golden, T.D., J. F. Veiga, and R. N. Dino 2008. The Impact of Professional Isolation on Teleworker Job Performance and Turnover Intentions: Does Time Spent Teleworking, Interacting Face-toFace, or Having Access to Communication-Enhancing Technology Matter? *Journal of Applied Psychology* 93 (6): 1412–1421.

Gouldner, A. W. 1960. The Norm of Reciprocity: A Preliminary Statement. *American Sociological Review* 25 (2): 161–178. https://doi.org/10.2307/2092623.

Greer, T.W., and S. C. Payne 2014. Overcoming Telework Challenges: Outcomes of Successful Telework Strategies. *The Psychologist-Manager Journal* 17 (2): 87–111. https://psycnet.apa.org/doi/10.1037/mgr0000014.

Guerin, B. 1993. Social Facilitation. Cambridge University Press.

Guggenmos, R.D., M. D. Piercey, and C. P. Agoglia 2018. Custom Contrast Testing: Current Trends and a New Approach. *The Accounting Review* 93 (5): 223–244. https://doi.org/10.2308/accr-52005.

Hannan, R.L., F. W. Rankin, and K. L. Towry 2006. The Effect of Information Systems on Honesty in Managerial Reporting: A Behavioral Perspective. *Contemporary Accounting Research* 23 (4): 885–918. https://doi.org/10.1506/8274-J871-2JTT-5210.

Hayes, A.F. 2018. Introduction to mediation, moderation, and conditional process analysis. The Guilford Press, New York, London.

Hendriks, A. 2012. SoPHIE - Software Platform for Human Interaction Experiments. *Working Paper*.

Houser, D., S. Vetter, and J. Winter 2012. Fairness and cheating. *European Economic Review* 56 (8): 1645–1655. https://doi.org/10.1016/j.euroecorev.2012.08.001.

Jensen, M.C., and W. H. Meckling 1976. Theory of the firm: managerial behavior, agency costs and ownership structure. *Journal of Financial Economics* 3 (4): 305–360. https://doi.org/10.1016/0304-405X(76)90026-X.

Johnson, N.D., and A. A. Mislin 2011. Trust games: A meta-analysis. *Journal of Economic Psychology* 32 (5): 865–889. https://doi.org/10.1016/j.joep.2011.05.007.

Kaiser, H.F. 1970. A second generation little jiffy. *Psychometrika* 35 (4): 401–415. https://doi.org/10.1007/BF02291817.

Kelly, K., E. LaMothe, and L. Baudot 2024. Remote “Helicopter Bosses”: Employee Perceptions of the Effects of Supervisory Controls and Remote Work During the COVID - 19 Pandemic*. *Accounting Perspectives* 23 (3): 333–373.

Ko, Y.H., and I. G. Baek 2024. The Effect of Computer Monitoring on Employees’ Productivity in Telecommuting Arrangements. *Management Science*: 1–20.

Koch, C., and S. E. Salterio 2017. The Effects of Auditor Affinity for Client and Perceived Client Pressure on Auditor Proposed Adjustments. *The Accounting Review* 92 (5): 117–142. https://www.jstor.org/stable/26551266.

Kroher, M., and T. Wolbring 2015. Social control, social learning, and cheating: Evidence from lab and online experiments on dishonesty. *Social science research* 53: 311–324. https://doi.org/10.1016/j.ssresearch.2015.06.003.

Lerner, J.S., and P. E. Tetlock 1999. Accounting for the Effects of Accountability. *Psychological Bulletin* 125 (2): 255–275. https://doi.org/10.1037/0033-2909.125.2.255.

Mazar, N., O. Amir, and D. Ariely 2008. The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research* 45 (6): 633–644. https://doi.org/10.1509/jmkr.45.6.633.

McKinsey & Company 2022. Americans are embracing flexible work—and they want more of it. https://www.mckinsey.com/industries/real-estate/our-insights/americans-are-embracing-flexible-work-and-they-want-more-of-it#/.

Miller, M.K., R. L. Mandryk, M. V. Birk, A. E. Depping, and T. Patel 2017. Through the Looking Glass: Effects of Feedback on Self-Awareness and Conversation during Video Chat. *Proceedings of the 2017 CHI conference on human factors in computing systems*: 5271–5283. https://doi.org/10.1145/3025453.3025548.

Osman, Y. 2024. Zoff bei Deutscher Bank – Verdi startet Petition. *Handelsblatt*. https://www.handelsblatt.com/finanzen/banken-versicherungen/banken/homeoffice-zoff-bei-deutscher-bank/100089278.html.

Owl Labs 2022. State of Remote Work 2022. https://resources.owllabs.com/hubfs/2022-10_SORW/Owl-Labs-State-of-Remote-Work-2022.pdf.

Pelletier, L.G., M. S. Fortier, R. J. Vallerland, and N. M. Briere 2001. Associations Among Perceived Autonomy Support, Forms of Self-Regulation, and Persistence: A Prospective Study. *Motivation and emotion* 25 (4): 279–306. https://doi.org/10.1023/A:1014805132406.

Piacenza, J., E. Moquin, L. Maxwell, A. He, L. Roeschke, M. Lynott, C. Principato, L. Whalen, R. Graziano, C. Tassin, and M. Howe 2022. New Workers, New Normal: Future of Work Report 2022. https://pro.morningconsult.com/analyst-reports/future-of-work-2022.

Plant, R.W., and R. M. Ryan 1985. Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of Personality* 53 (3): 435–449. https://doi.org/10.1111/j.1467-6494.1985.tb00375.x.

Porter, J. 2022. Elon Musk reportedly declares remote work 'no longer acceptable' at Tesla. *The Verge*. https://www.theverge.com/2022/6/1/23149784/elon-musk-tesla-remote-work-leaked-email-40-hours.

Rankin, F.W., S. T. Schwartz, and R. A. Young 2008. The Effect of Honesty and Superior Authority on Budget Proposals. *The Accounting Review* 83 (4): 1083–1099. https://doi.org/10.2308/accr.2008.83.4.1083.

Ravid, D.M., D. L. Tomczak, J. C. White, and T. S. Behrend 2020. EPM 20/20: A Review, Framework, and Research Agenda for Electronic Performance Monitoring. *Journal of Management* 46 (1): 100–126. https://doi.org/10.1177/0149206319869435.

Reid, E. 2015. Embracing, passing, revealing, and the ideal worker image: How people navigate expected and experienced professional identities. *Organization Science* 26 (4): 997–1017. https://doi.org/10.1287/orsc.2015.0975.

Ross, S.A. 1973. The Economic Theory of Agency: The Principal's Problem. *American Economic Association* 63 (2): 134–139.

Rupietta, K., and M. Beckmann 2018. Working from home: What is the effect on employees' effort? *Schmalenbach Business Review* 70 (1): 25–55. https://doi.org/10.1007/s41464-017-0043-x.

Ryan, R.M., and E. L. Deci 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55 (1): 68–78. https://psycnet.apa.org/doi/10.1037/0003-066X.55.1.68.

Schedlinksy, I., M. Schmidt, and A. Wöhrmann 2020. Interaction of information and control systems: How the perception of behavior control affects the motivational effect of relative performance information. *Accounting, Organizations and Society* 86. https://doi.org/10.1016/j.aos.2020.101171.

Schlund, R., and E. M. Zitek 2024. Algorithmic versus human surveillance leads to lower perceptions of autonomy and increased resistance. *Communications Psychology* 2 (1). https://doi.org/10.1038/s44271-024-00102-8.

Shu, L.L., N. Mazar, F. Gino, D. Ariely, and M. H. Bazerman 2012. Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences of the United States of America* 109 (38): 15197–15200. https://doi.org/10.1073/pnas.1209746109.

Silvia, P.J., O'Brien, and Maureen E. 2004. Self-Awareness and Constructive Functioning: Revisiting “the Human Dilemma”. *Journal of Social & Clinical Psychology* 23 (4): 475–489. https://doi.org/10.1521/jscp.23.4.475.40307.

Sprinkle, G.B. 2003. Perspectives on experimental research in managerial accounting. *Accounting, Organizations and Society* 28 (2-3): 287–318. https://doi.org/10.1016/S0361-3682(01)00058-7.

Stewart, D.W. 1981. The Application and Misapplication of Factor Analysis in Marketing Research. *Journal of Marketing Research* 18 (1): 51–62. https://doi.org/10.1177/002224378101800105.

Tan, H.-T., and A. Kao 1999. Accountability Effects on Auditors' Performance: The Influence of Knowledge, Problem-Solving Ability, and Task Complexity. *Journal of Accounting Research* 37 (1): 209.

Tessier, S., and D. Otley 2012. A conceptual development of Simons' Levers of Control framework. *Management Accounting Research* 23 (3): 171–185. https://doi.org/10.1016/j.mar.2012.04.003.

Waterson, J. 2016. Daily Telegraph Installs Workplace Monitors On Journalists' Desks. *BuzzFeed*. https://www.buzzfeed.com/jimwaterson/telegraph-workplace-sensors.

Webb, R.A. 2002. The impact of reputation and variance investigations on the creation of budget slack. *Accounting, Organizations and Society* 27 (4-5): 361–378. https://doi.org/10.1016/S0361-3682(01)00034-4.

Wicklund, R.A. 1975. Objective Self-Awareness. *Journal of Experimental Social Psychology* 8: 233–275. https://doi.org/10.1016/S0065-2601(08)60252-X.

Wicklund, R.A., and S. Duval 1971. Opinion change and performance facilitation as a result of objective self-awareness. *Journal of Experimental Social Psychology* 7 (3): 319–342. https://doi.org/10.1016/0022-1031(71)90032-1.

TABLE 1

Descriptive Statistics (mean, [standard deviation])

	Location[a]						**Total**
	Office			**Telework**			
	Monitoring[b]			**Monitoring**[b]			
	present	absent	**Total**	present	absent	**Total**	
Number of Subjects	21	20	**41**	24	23	**47**	88
Panel A: Effort (n = 88)							
Number of sliders moved[c]	228.43	266.20	**246.85**	329.21	338.96	**333.98**	293.39
	[80.83]	[119.85]	**[102.25]**	[97.47]	[92.50]	**[94.16]**	[106.80]
Panel B: Misreporting (n = 88)							
Extent of Misreporting[d]	0.51	0.68	**0.59**	0.49	0.53	**0.51**	0.55
	[0.31]	[0.26]	**[0.29]**	[0.26]	[0.30]	**[0.28]**	[0.29]

[a] *Location* has two levels: office and telework. Participants either conducted the experiment in the experimental laboratory at the university (*office*) or at home (*telework*). Location is coded as 1 for the telework condition and 0 otherwise.

[b] *Monitoring* has two levels: present or absent. Participants were either monitored (*present*) or not (*absent*). Monitoring is coded as 1 when monitoring is present and 0 otherwise.

[c] *Number of sliders moved* captures the total adjustments made to sliders during the effort task, regardless of their correctness.

[d] *Extent of Misreporting* represents the extent to which participants lied relative to how much they could have lied. The variable takes on values between 0 and 1. Values closer to 1 indicate more dishonest behavior.

TABLE 2

Effects of Location and Monitoring on Effort – H1a & H2a

Dependent variable: Effort[a] (n = 88)

Panel A: ANCOVA for H1a

Source of variation	SS	df	MS	F-ratio	p-value[f]
Model	259332.40	10	25933.24	2.72	< 0.01
Location[b]	162931.29	1	162931.29	17.12	< 0.01
Monitoring[c]	17250.67	1	17250.67	1.81	0.18
Location x Monitoring	5365.95	1	5365.95	0.56	0.46
Device[d]	3734.79	1	3734.79	0.39	0.53
Location Preference[e]	68588.57	6	11431.43	1.20	0.32
Error	732946.86	77	9518.79		

Panel B: Planned Contrasts for H2a

Source	SS	df	MS	F-ratio	p-value[f]
Model custom contrast[g]	183426.99	1	183426.99	19.27	< 0.01
Error	732946.46	77	9518.79		

[a] The dependent variable *Effort* is operationalized through the number of sliders moved.

[b] *Location* has two levels: office and telework. Participants either conducted the experiment in the experimental laboratory at the university (*office*) or at home (*telework*). Location is coded as 1 for the telework condition and 0 otherwise.

[c] *Monitoring* has two levels: present or absent. Participants were either monitored (*present*) or not (*absent*). Monitoring is coded as 1 when monitoring is present and 0 otherwise.

[d] *Device* is a dummy variable capturing the usage of a mouse (1) or touchpad (0) during the slider task.

[e] *Location Preference* is measured in the PEQ through a 7-point Likert scale asking participants to what extent they agree with the statement: "I prefer to complete work assignments at home rather than at the university." (7- point Likert scale, reversed).

[f] All p-values are two-tailed.

[g] The contrast coefficients are -7 for *Office/Monitoring present*, -1 for *Office/Monitoring absent*, +3 for *Telework/Monitoring present*, and +5 for *Telework/Monitoring absent*.

TABLE 3

Effects of Location and Monitoring on Misreporting – H1b & H2b

Dependent variable: Misreporting[a] (n = 88)

Panel A: ANCOVA for H1b

Source of variation	**SS**	**df**	**MS**	**F-ratio**	**p-value[e]**
Model	1.34	9	0.15	2.03	0.05
Location[b]	0.30	1	0.30	4.02	0.05
Monitoring[c]	0.24	1	0.24	3.32	0.07
Location x Monitoring	0.10	1	0.10	1.32	0.25
Preference Location[d]	0.85	6	0.14	1.94	0.09
Residual	5.75	78	0.07		
Total	7.09	87	0.08		

Panel B: Planned Contrasts for H2b

Source	**SS**	**df**	**MS**	**F-ratio**	**p-value[e]**
Model custom contrast[f]	0.59	1	0.59	7.99	< 0.01
Error	5.75	78	0.07		

[a] The dependent variable *Misreporting* takes on values between 0 and 1. Values close to 0 indicate honest behavior. The variable is calculated by the sum of the difference between reported cost – actual cost divided by the maximum project revenue of 2,000 – actual cost.

[b] *Location* has two levels: office and telework. Participants either attended the experiment in the experimental laboratory at the university (*office*) or at home (*telework*). Location is coded as 1 for the telework condition and 0 otherwise.

[c] *Monitoring* has two levels: present or absent. Participants were either monitored (*present*) or not (*absent*). Monitoring is coded as 1 when monitoring is present and 0 otherwise.

[d] *Location Preference* is measured in the PEQ through a 7-point Likert scale asking participants to what extent they agree with the statement: "I prefer to complete work assignments at home rather than at the university." (7- point Likert scale, reversed).

[e] All p-values are two-tailed.

[f] The contrast coefficients are -1 for *Office/Monitoring present*, +4 for *Office/Monitoring absent*, -2 for *Telework/Monitoring present*, and -1 for *Telework/Monitoring absent*.

TABLE 4

Items from the PEQ (7-points scale) used for the Path Analyses

Panel A: Autonomy Item

I felt restricted in my actions by my employer's decisions regarding the organization of my work.[a]
(endpoints: totally agree and totally disagree)

Panel B: Factor Analyses for Motivation and Reciprocity

Motivation Items

I was motivated during the experiment.[a]
(endpoints: totally agree and totally disagree)

It was important for me to master the task well.[a]
(endpoints: totally agree and totally disagree)

Reciprocity Items	Used for **Effort reciprocity**	Used for **Misreporting reciprocity**
General Reciprocity Item		
I retaliated against my employer's decision regarding my place of work through my decisions in the tasks. *(endpoints: totally agree and totally disagree)*	x	x
I was annoyed by my employer's decision regarding my place of work. *(endpoints: totally agree and totally disagree)*	x	x
Effort-/Misreporting-specific Reciprocity Items		
I wanted to thank my employer for the choice of my working conditions (place of work and monitoring) and therefore moved a lot of sliders.[a] *(endpoints: totally agree and totally disagree)*	x	
I wanted to avoid that my employer received a lower payout for the project task because of me.[a] *(endpoints: totally agree and totally disagree)*		x

[a] Marked items have been reversed for the respective factor calculation.

FIGURE 1

Experimental Procedure

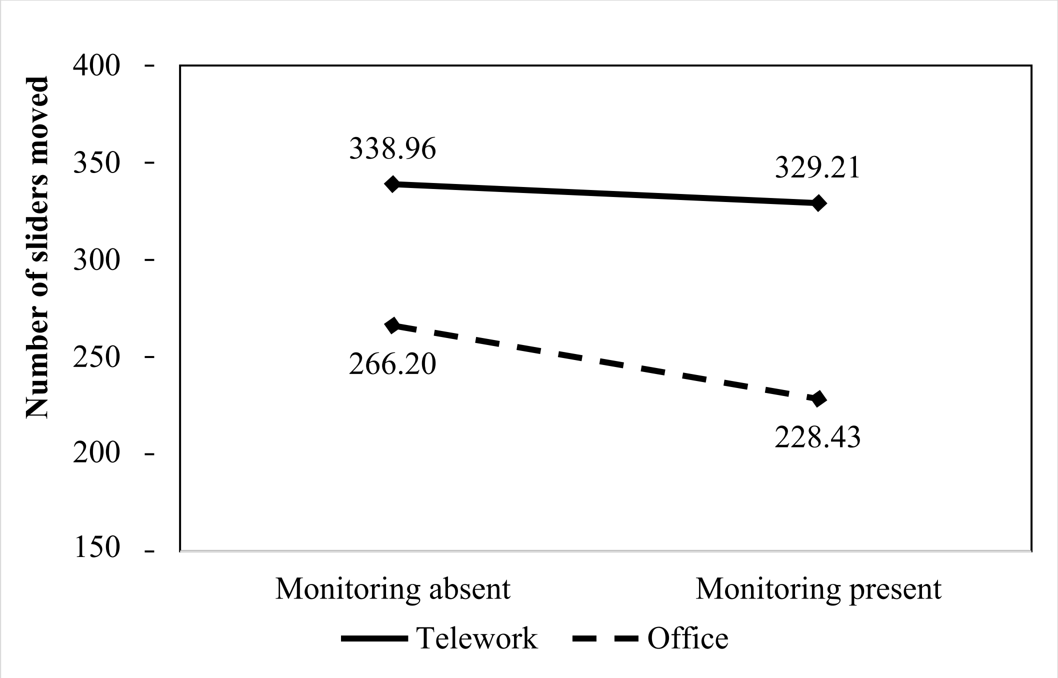
Pre-Study Survey	Decision about the working environment
Main Experiment Part 1 Quiz	Introduction to the experiment and comprehension quiz
Part 2 Start Monitoring	Joining the online video conferencing room[a]
Part 3 First Task	Effort Task[b] for A-Split Reporting Task[c] for B-Split
Part 4 Second Task	Effort Task[b] for B-Split Reporting Task[c] for A-Split
Part 5 End Monitoring	Leaving the online video conferencing room[a]
Part 6 PEQ	Post-experimental questionnaire

[a] Applies only to participants in the monitoring conditions.
[b] The effort task (slider task) consists of solving as many sliders as possible with the option to play up to 10 rounds tic-tac-toe against the computer.
[c] The reporting task consists of four cost reports. Participants learn the actual costs and have to decide which costs they report.

FIGURE 2

Observed Effects of Location and Monitoring on Effort and Misreporting

Panel A: Observed Effects on Effort (n = 88), H1a and H2a



Panel B: Observed Effects on Misreporting (n = 88), H1b and H2b

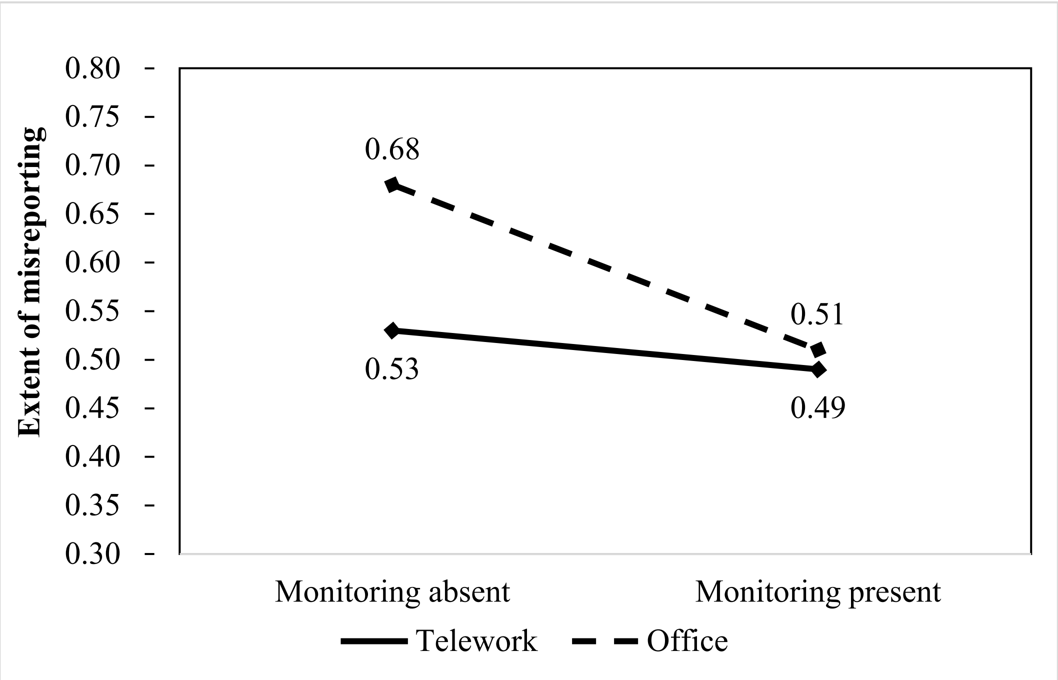
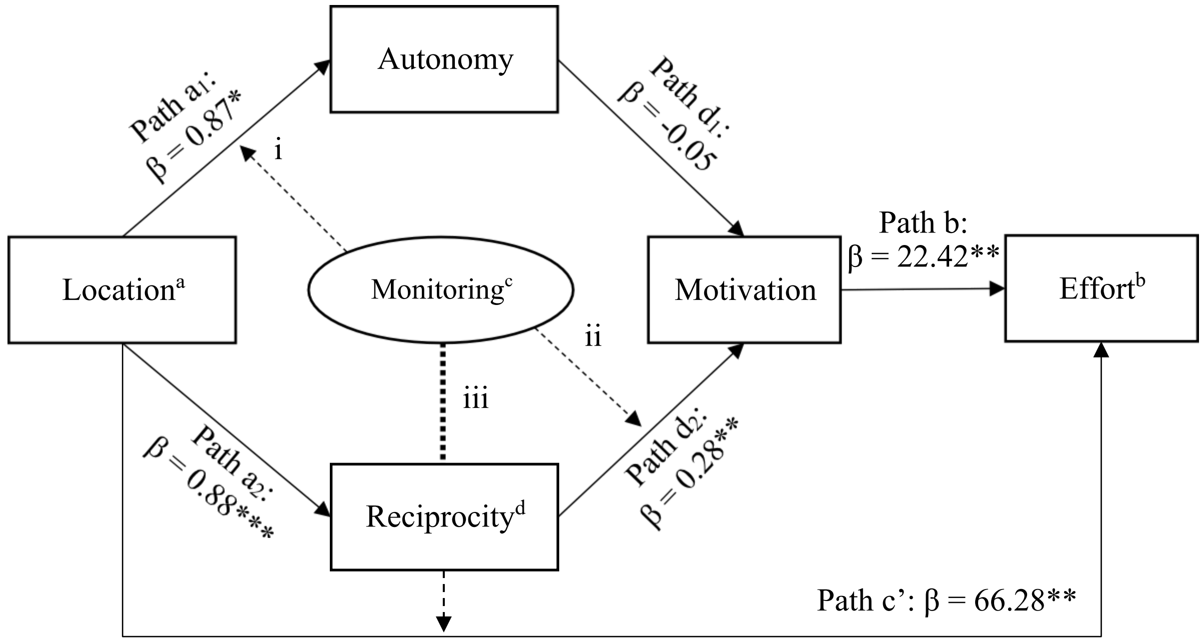


FIGURE 3
Path Analysis

Panel A: Effort



90% BCI for the indirect effect through reciprocity and motivation (a×b; monitoring absent): **(0.28, 17.15)**

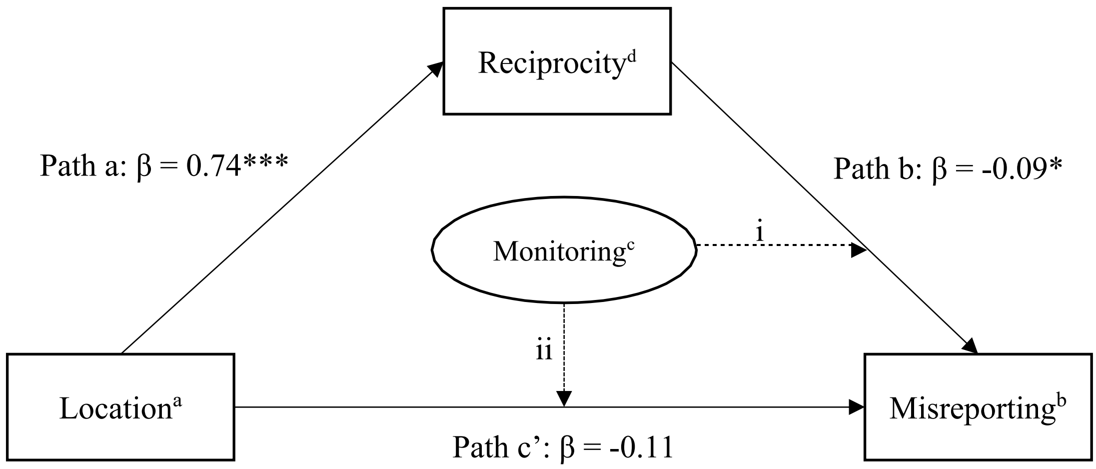
90% BCI for the indirect effect through reciprocity (a×b; monitoring present): **(2.17, 33.49)**

90% BCI for the difference in the indirect effects through reciprocity: **(1.28, 24.43)**

If zero does not appear within the 90% bias-corrected confidence-interval (BCI), then the indirect effect explained by reciprocity (a×b) has statistical significance at the 10% level (displayed in **bold**).

Place x Monitoring ***i***: -1.67, $t = -2.66$, $p < 0.01$; direct effect of monitoring on autonomy: -1.30, $t = -2.85$, $p < 0.01$.
Reciprocity x Monitoring ***ii***: 0.41, $t = 2.16$, $p = 0.03$; direct effect of monitoring on motivation: -0.37, $t = -1.63$, $p = 0.11$.
Place x Monitoring ***iii***: 29.19, $t = 0.70$, $p = 0.49$; direct effect of monitoring on effort: -31.06, $t = -1.02$, $p = 0.31$.
[a] *Location* is the independent variable and coded as 1 for the telework condition and 0 otherwise.
[b] *Effort* is the dependent variable measured by the number of sliders moved.
[c] *Monitoring* is the moderator and coded as 1 when monitoring is present and 0 otherwise.
[d] The *Reciprocity* factor for effort consists of three items from the PEQ (see Table 4).
Panel A depicts our path model estimated using the PROCESS macro in SPSS (Hayes 2018). **Path a_1** represents the effect of location on autonomy ("I felt restricted in my actions by my employer's decisions regarding the organization of my work." (7-point Likert scale, reversed). Higher values indicate a higher perceived autonomy). **Path a_2** reflects the effect of location on (effort) reciprocity (a factor composed of the three questions "I wanted to thank my employer for the choice of my working conditions (place of work and monitoring) and therefore moved a lot of sliders." (7-point Likert scale, reversed), "I retaliated against my employer's decision regarding my place of work through my decisions in the tasks." (7-point Likert scale) and "I was annoyed by my employer's decisions regarding my place of work." (7-point Likert scale). Higher values indicate higher reciprocal behavior.) **Path b** denotes the effect of motivation on effort (motivation is a factor composed of the two questions "I was motivated during the experiment" and "It was important for me to master the task well." (7-point Likert scale, reversed). Higher values indicate higher motivation.). **Path c'** represents the direct effect of location on effort. **Path d_1** reflects the effect of autonomy on motivation. **Path d_2** represents the effect of (effort) reciprocity on motivation.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

Panel B: Misreporting



90% BCI for the indirect effect through reciprocity (a×b, monitoring absent): **(-0.13, -0.01)**

90% BCI for the indirect effect through reciprocity (a×b, monitoring present): **(-0.14, 0.00)**

90 % BCI for the difference in the indirect effects (-0.08, 0.09)

If zero does not appear within the 90% bias-corrected confidence-interval (BCI), then the indirect effect explained by reciprocity (a×b) has statistical significance at the 10% level (displayed in **bold**).

*Monitoring x Reciprocity **i***: 0.02, t = 0.32, p = 0.75.
*Monitoring x Place **ii***: 0.11, t = 0.91, p = 0.36; direct effect of monitoring on misreporting: -0.18, t = -2.09, p = 0.04.
[a] *Location* is the independent variable and coded as 1 for the telework condition and 0 otherwise.
[b] *Misreporting* is the dependent variable measured by the sum of reported cost – actual cost divided by the sum of the maximum project revenue of 2,000 – actual cost over all four projects.
[c] *Monitoring* is the moderator and coded as 1 when monitoring is present and 0 otherwise.
[d] The *Reciprocity* factor for misreporting consists of three items from the PEQ (see Table 4).
Panel B depicts our moderated mediation model estimated using the PROCESS macro in SPSS (Hayes 2018). Path **a** represents the effect of location on (misreporting) reciprocity (a factor composed of the three questions "I wanted to avoid that my employer received a lower payout for the project task because of me." (7-point Likert scale, reversed), "I retaliated against my employer's decision regarding my place of work through my decisions in the tasks." (7-point Likert scale, reversed) and "I was annoyed by my employer's decisions regarding my place of work." (7-point Likert scale). Higher values indicate higher reciprocal behavior.). Path **b** reflects the effect of (misreporting) reciprocity on misreporting. Path **c'** represents the direct effect of location on misreporting.
*** p < 0.01, ** p < 0.05, * p < 0.10

Study no.	**3**			
Title	**Debiasing Training—Mitigating Cognitive Biases Via Training Interventions: A Systematic Literature Review**			
Authors	Rebecca Sabel, Arnt Wöhrmann, Marlene Alt			
Author contribution		Sabel	Wöhrmann	Alt
	Numeric share	*0.9*	*0.05*	*0.05*
	Conceptual development of research question	✓	✓	
	Development of theory	✓		✓
	Methodology	✓		
	Acquisition of data	✓		✓
	Analysis/interpretation of data	✓		
	Writing the manuscript	✓	✓	
Publication status	Revise & Resubmit (R&R) *Accounting Perspectives (VHB-Rating 2024: B)*			
Research approach	Literature review			
Language	English			
Abstract	This paper presents a systematic review of research on the effectiveness of debiasing training interventions. Specifically, we identify six training approaches and evaluate their impact on reducing motivational, statistical, and informational biases. Our findings reveal that serious games have been the most frequently studied and appear to offer the greatest potential for debiasing despite their increased complexity. We also identify moderators that have been examined for their potential influence on training effectiveness, though current findings suggest a limited impact on bias mitigation. Additionally, we explore potential mediating factors that may explain why some training approaches outperform others. Finally, we discuss practical implications for implementing debiasing training and suggest directions for future research.			

Debiasing Training—Mitigating Cognitive Biases Via Training Interventions: A Systematic Literature Review

Rebecca Sabel
Justus Liebig University Giessen (Germany)
Rebecca.Sabel@wirtschaft.uni-giessen.de
Licher Str. 62,
35394 Giessen, Germany

Arnt Wöhrmann
Justus Liebig University Giessen (Germany)
Arnt.Woehrmann@wirtschaft.uni-giessen.de

Marlene Alt
Justus Liebig University Giessen (Germany)
Marlene.S.Alt@wirtschaft.uni-giessen.de

Abstract

This paper presents a systematic review of research on the effectiveness of debiasing training interventions. Specifically, we identify six training approaches and evaluate their impact on reducing motivational, statistical, and informational biases. Our findings reveal that serious games have been the most frequently studied and appear to offer the greatest potential for debiasing despite their increased complexity. We also identify moderators that have been examined for their potential influence on training effectiveness, though current findings suggest a limited impact on bias mitigation. Additionally, we explore potential mediating factors that may explain why some training approaches outperform others. Finally, we discuss practical implications for implementing debiasing training and suggest directions for future research.

Keywords: Cognitive Biases, Debiasing, Management Accounting, Serious Games, Training

JEL: M41, M10

I. Introduction

Decision-making, whether conscious or unconscious, is central to managers' daily responsibilities and forms the basis for firm success. However, if success does not materialize, cognitive biases, i.e., psychologically induced deviations from economically rational decision-making, are often to blame (Soll et al., 2015; Tversky & Kahneman, 1974). Cognitive biases are costly and can negatively affect firm performance. For instance, the anchoring bias—the tendency to rely on initial information rather than integrating new insights (Tversky & Kahneman, 1974)—can reduce shareholder returns by 1.5% to 3.9% during capital allocation (Baer et al., 2017). Further, overconfident managers often overestimate benefits and underestimate costs for high-risk projects, a tendency evidenced by the numerous cases of CEOs overpaying for acquisition targets (Malmendier & Tate, 2008). Hence, biases can lead to inefficient capital allocation at best and, at worst, to massive losses or even bankruptcy (Goto, 2007).

A survey by Kreilkamp et al. (2021) highlights that while 50% of firms *occasionally* encounter all surveyed biases and three of these even *frequently*, only 12% use systematic debiasing strategies. This gap underscores the need for effective debiasing, i.e., techniques that mitigate cognitive biases and thus support rational decision-making.

As business partners, management accountants play a key role in ensuring rational decision- making (Goretzki et al., 2013; Järvenpää, 2007). Beyond unbiased information collection and reporting, management accountants can actively shape a more rational decision-making environment. First, they serve as independent, trusted counterparts who challenge bias-driven assumptions in analyses or discussions (Schäffer, 2013). Second, they should identify and recognize cognitive biases and design management controls that mitigate rather than reinforce them (Ding & Beaulieu, 2011). This may involve adjusting specific controls within the management control system and fostering a corporate culture that embraces debiasing

(Kreilkamp et al., 2021). Most importantly, management accountants must embed debiasing into decision-making by developing expertise and educating decision-makers on its relevance and practical application (Kreilkamp et al., 2021). The latter, in the form of debiasing training, has gained attention as a relatively unexplored research area alongside well-researched techniques such as incentives, accountability, and nudges (Scopelliti, 2023).

In debiasing trainings, decision-makers are systematically taught on how to identify and mitigate one or more cognitive biases. Notable examples include initiatives by Facebook and Google, which use videos and workshops to address biases primarily related to race and gender, in hiring and performance evaluations (Google, 2013; Sandberg, 2015). While these video-based trainings are simple to implement, researchers have increasingly explored more complex interventions and also other, more general biases in decision-making. Another prominent training approach, supported by a U.S. government research program, is the use of serious games for debiasing. Unlike traditional video games, serious games are designed for educational purposes They present bias-prone scenarios where players engage in interactive quizzes and learn strategies to recognize and mitigate cognitive biases (Scopelliti, 2023).

Responding to Aczel et al.'s (2015) call for systematic research, this study reviews the effectiveness, design choices, and mediating factors of various forms of debiasing training encompassing but not limited to serious games and video trainings. Our objectives are therefore threefold: In a first step, we outline the theoretical mechanisms through which debiasing training can impact cognitive processing. We categorize training interventions and assess their effectiveness. We synthesize findings by comparing the frequency of positive versus null effects reported across studies, highlighting the most effective interventions where data allows such conclusions. Additionally, we examine the persistence of training effects and their transferability to untrained contexts.

Second, we identify and categorize design elements of the two most prevalent training interventions (i.e., serious game and awareness training), assessing their role in reducing susceptibility to cognitive biases. We also highlight success factors and discuss their impact. Third, we explore mediating effects that explain why some trainings are more effective than others. We add graphical summaries of each training's mitigating effects, moderating design elements, and potential mediating factors. Finally, we discuss future research directions.

This review contributes to both theory and practice by providing a broader perspective on debiasing than existing research, which often focuses on specific cognitive biases or isolated debiasing approaches (e.g., Brüggen & Luft, 2016; Loh et al., 2019). For instance, Berthet (2021a, 2021b) offers comprehensive reviews of cognitive biases but does not explore debiasing strategies, while Jugnandan and Willows (2023) focus on debiasing but limit their analysis to financial decision-making. Korteling et al. (2021) systematically review bias mitigation interventions but restrict their scope to retention and transfer effects, analyzing twelve studies with eleven on game- or video-based interventions. In contrast, our review systematically examines debiasing-related trainings across all empirical research fields, providing a comprehensive overview of various interventions and their effectiveness. We identify six training types: serious games, training videos, reading serious game scripts, observational learning, awareness training, and analogical training. We find that serious games appear the most promising. Yet, simple-to-implement-trainings, such as video and awareness trainings, also have notable effects but more research is necessary to recommend these types for broad usage.

Further, we extend previous literature by examining key design choices and provide recommendations for effective training design in practice. For serious games, we investigate several game characteristics and find that at least some studies show a positive effect for longer duration, explicit training elements (e.g., quizzes), and repeated gameplay. Interestingly, however, the majority of moderators (player type, feedback timing, game style, and rewards) does not have an effect. For awareness training, both short and long decision times yield debiasing effects, while repetition does not enhance bias mitigation.

From a theory perspective, we contribute by identifying mediators for the effectiveness of debiasing trainings that explain why some trainings are more effective than others. For instance, results show that playing a serious game is more effective than watching a training video (e.g., Morewedge et al., 2015; Symborski et al., 2014; Veinott et al., 2013). This appears to be due to greater engagement and the ability to learn from mistakes, which enhance players' self-determination, motivation, and enjoyment—ultimately facilitating deep learning. We use these findings to derive general recommendations for firms to implement debiasing trainings. For example, firms should select training based on relevant biases and employees' motivation to engage in debiasing. When motivation is low, interactive interventions like serious games appear more effective. If motivation is high, simpler and less costly options, such as awareness or video training, may suffice.

Finally, our review highlights several avenues for future research. First, future research should investigate the longevity of debiasing effects, examining whether these effects persist beyond a couple of weeks. Additionally, research should focus on the transferability of debiasing outcomes to real workplace environments to ensure practical benefits. Further, some training methods, such as reading serious game scripts, observational training, and analogical training, have been barely investigated; expanding research on these methods and exploring their application to additional biases is essential. Lastly, mediating effects of training interventions remain underexplored and warrant further examination.

The remainder of this paper proceeds as follows. Section 2 elaborates on the theoretical background of biases and describes existing training interventions. Section 3 presents the study selection process and our review. Section 4 discusses key insights, practical implications, and directions for future research. Finally, section 5 provides the conclusion.

II. Background

Cognitive biases

Cognitive biases are systematic deviations from economically optimal and rational decisions (Ehrlinger et al., 2016; Tversky & Kahneman, 1974). Biases contradict the *homo oeconomicus* model, which assumes decision-makers have complete information, stable preferences, and sufficient resources to make rational, error-free decisions that maximize expected utility (Hirsch, 2007; Kirchgässner, 2013). Behavioral research documents that decision maker's cognitive capacity for information processing is limited and influenced by uncertainty and emotions (Simon, 1978).

As outlined in Kahneman's (2011) dual-process theory, decision-makers rely more on intuitive System 1 thinking ("fast thinking") than on rational System 2 thinking ("slow thinking"). While System 1 is fast, intuitive and effortless, it makes use of heuristics. Heuristics are simplified decision rules that select or refine relevant information (i.e., rules of thumb), which allows decision-makers to cope with uncertainty, time pressure and information overload (Tversky & Kahneman, 1974). While heuristics can lead to satisfying decisions, they do not always result in optimal and rational outcomes—a concept known as "bounded rationality". This occurs because not all information is fully processed, and logical principles or probabilities may be overlooked (Kahneman & Tversky, 1972; Montibeller & Winterfeldt, 2015; Morewedge & Kahneman, 2010; Tversky & Kahneman, 1974).

As a result, System 1 heuristics often lead to cognitive biases. These biases can be mitigated by engaging System 2, which is rule-based and involves effortful, and logical thinking that supports rational decision-making (Kahneman, 2011). However, System 1 often dominates managerial decision-making, as it is intuitive and much faster than the deliberate System 2, which requires weighing all options and their potential outcomes. Therefore, strategies must be developed that activate System 2 thinking and reduce biases.

Research has identified a vast number of cognitive biases that occur consciously or unconsciously. These biases can be classified as *motivational biases*, i.e., motive-driven decisions, *statistical biases*, i.e., decisions based on misjudgments of statistical correlations, and *informational biases*, i.e., decisions based on environment-driven information selection and processing (Montibeller & Winterfeldt, 2015). Table 1 presents prominent examples of biases in these three groups that are addressed in the debasing studies reviewed in section 3.

[Insert Table 1 here]

Debiasing and debiasing training

Cognitive biases are robust and reduce decision-making quality, leading to detrimental effects on firm performance (Soll et al., 2015). This creates a need for systematic *debiasing*—a crucial task in management accounting. Debiasing refers to interventions—such as tools or frameworks—that aim at rationalizing judgment and decision-making by reducing or eliminating biases and their adverse effects (Fischhoff, 1982; Kreilkamp et al., 2019; Larrick, 2004; Sellier et al., 2019a; Soll et al., 2015). Thereby, debiasing improves the processing of available information without neglecting or violating normative rules and logic (Soll et al., 2015).

Debiasing trainings are designed to help decision-makers reduce cognitive biases and vary in format. Fischhoff (1982) identifies key elements of effective training: (1) raising awareness of biases and one's susceptibility, (2) explaining biases and their impact, (3) providing (personalized) feedback on bias susceptibility, and (4) offering practice-based training with feedback, coaching, and debiasing strategies.[59] Not every debiasing training in the literature includes all elements, but they consistently incorporate at least two, i.e., creating awareness and sensitizing individuals to their impact. The latter is crucial, as research shows that raising awareness alone does not suffice (e.g., Beaulac & Kenyon, 2018; Fischhoff, 1982;

[59] For instance, Morewedge et al. (2015) provide a list of debiasing strategies taught in their training.

Larrick, 2004; Weinstein & Klein, 1995), likely due to the deeply ingrained nature of many biases (Stanovich, 2011). Consequently, we define debiasing training as any educational program that, at a minimum, explains biases and includes strategies to enhance bias awareness. Below, we describe the six types of debiasing trainings that are discussed in the literature: (I) s*erious game*, (II) *training video*, (III) *reading serious game script*, (IV) *observational training*, *awareness training*, and (VI) *analogical training*.

(I) Serious games: Serious games are computer games intended to teach and educate the user rather than to purely entertain (Stapleton, 2004).[60] Originally, this intervention was designed to mitigate biases by US government intelligence analysts (Sellier et al., 2019a) and has been applied in the military and health care field (Stapleton, 2004). Prominent examples of such serious games are *Missing, Heuristica, Cycles, Enemy of Reason,* and *Macbeth.*[61]

Serious games usually consist of two phases: a learning phase and an action phase. In the learning phase, cognitive biases are elicited from the players during game play and are subsequently defined and explained. The training consists of quizzes and scenarios that expose players to situations where biases are likely to arise. According to the players' actions and answers, the training provides individual feedback on the player's susceptibility to biases (Sellier et al., 2019a). Lastly, the training introduces strategies to mitigate or prevent cognitive biases, such as considering alternative explanations (i.e., consider-the-opposite strategy), evaluating possible outcomes (i.e., premortem analysis), adjusting anchors, considering statistical principles and accounting for counterevidence (Morewedge et al., 2015). In the action phase, players apply the new knowledge to related and novel problems (Bush, 2017;

[60] For a detailed description of the development process of serious games in the debiasing context see Mullinix et al. (2013) and Symborski et al. (2017).

[61] These games were developed as part of the Sirius program, a research initiative aimed at enhancing decision making and reducing cognitive biases. The program was launched under the oversight of the Intelligence Advanced Research Projects Activity (IARPA), a U.S. government agency that funds high-risk, high-reward research to improve intelligence analysis (Rhodes et al., 2017). For an analysis of the five mentioned serious games see Rhodes et al. (2017).

Morewedge et al., 2015; Mullinix et al., 2013; Symborski et al., 2017). The effectiveness of serious games in reducing biases is measured using pre- and post-training questionnaires that assess the susceptibility of committing cognitive biases. Some studies also include a follow-up test four to twelve weeks after the training intervention to evaluate long-term debiasing effects (Morewedge et al., 2015).

(II) Training video: Training videos are educational videos that use vignettes to illustrate real-world examples where people demonstrate cognitive biases.[62] An expert explains each bias, discusses its impact, and provides mitigation strategies (Clegg et al., 2014; Clegg et al., 2015; Morewedge et al., 2015). All training videos ensure that the core information on bias definitions and mitigation strategies aligns with the training they are compared to, typically serious games.

(III) Reading serious game script: This training approach uses a written transcript of the scenarios used in serious games. The script includes identical content to the serious game, with the same dialogues, descriptions, and environmental details (Poos et al., 2017). However, it lacks the game's visual richness and interactive elements. Instead of making decisions to navigate through specific scenarios, participants read through all possible options and outcomes.

(IV) Observational training: Observational trainings also build on serious games. However, instead of playing the game or reading the script, individuals observe others playing the game (Yoon et al., 2021). Thus, the observer receives the same examples, explanations, and debiasing strategies as the players, but without actively participating in gameplay.

(V) Awareness training: This approach delivers training content via text (e.g., educational passages) or verbal explanations (e.g., slide shows, lectures). For instance, participants receive text-based explanations on specific biases, often paired with practice

[62] Similar to the serious games, most training videos were developed by the Intelligence Advanced Research Projects Activity (2012, 2013) (IARPA I and II).

sessions and feedback (e.g., van Brussel et al., 2021). Similar methods include lecture-based sessions with slides (Aczel et al., 2015; Lee et al., 2016). Some awareness trainings are even simpler, providing only educational descriptions of biases without debiasing strategies or practice (e.g., AlKhars et al., 2019; Luong & Butler, 2022). Another method presents bias-prone questions, followed by written explanations of both the correct and biased responses (Boissin et al., 2022).

(VI) Analogical training: In analogical trainings, individuals learn about biases and debiasing strategies by identifying structural similarities between superficially different situations and associating these analogous situations with each other (Aczel et al., 2015). For instance, participants receive descriptions of typical bias-prone scenarios and are tasked with recognizing parallels between these scenarios to identify the underlying biases. Next, the biases are defined and explained. Participants then reflect on personal experiences with similar biases, reinforcing awareness. Finally, the training introduces strategies to prevent biases.

III. Review

Study selection process

Our systematic literature review follows the identification process suggested by Tranfield et al. (2003) and Hardies et al. (2024). First, we define eligibility criteria, including published studies on debiasing trainings. Second, we implement our search strategy by scanning *The Web of Science*, *Google Scholar*, *Science Direct,* and *EBSCO Discovery Service* using the search terms *(cognitive) biases, debiasing, training, intervention,* and *mitigation*.[63] Relevant studies are identified by reviewing titles and abstracts for the specific keywords and alignment with our research question. Additionally, we conduct a manual search by reviewing references of identified studies and use *The Web of Science* to find papers citing previously selected

[63] We explicitly exclude the keywords *modification, psychiatric, depression, race,* and *gender* to neglect the intensively investigated psychotherapeutic topic of bias modification training. A bias modification training is a depression intervention technique to alleviate depressive symptoms, for a literature review on these trainings see Li et al. (2023).

studies. Next, we merge identified studies from all databases and remove duplicates. Finally, we exclude studies lacking theoretical reasoning, do not sensitize for a specific bias, fail to investigate bias mitigation, or do not align with our definition of debiasing training.[64]

Our selection process leads to a total of 30 studies that address one or more debiasing trainings. The majority of these studies examine serious games (18) and/or training videos (13), followed by awareness training (12). Only one study each investigates the effect of reading a serious game script, observational and analogical training. Table 2 depicts the studies included in this review and the training approach used.

In the next section, we first assess the effectiveness of these training approaches in reducing biases. Next, we examine their impact on bias recognition and discrimination. We then analyze the role of moderators in serious games and awareness trainings. Finally, we explore potential mediators that may explain the (lack of) effectiveness of these interventions.

[Insert Table 2 here]

Empirical findings

Bias mitigation

Motivational Biases

For our discussion of the effectiveness of debasing training, the biases examined in the 30 papers are clustered into motivational, statistical, and informational biases. This approach seems useful, as certain training approaches may be more effective for specific bias types. Table 3 summarizes our findings.

The motivational biases examined are confirmation bias, fundamental attribution error, blind spot bias, and projection bias. Out of the 20 studies on motivational biases, 17 investigate the effect of serious games and/or training videos, often using training videos as a benchmark.

[64] Based on these criteria we exclude studies by Chang et al. (2016); Graf et al. (2012); Hutchinson et al. (2010); Jensen et al. (2016); Kovic (2019); Kucera (2020); Legaki et al. (2021); Mersch et al. (2013); Murata (2016); Neilens et al. (2009); Renner and Renner (2001); and Rumeser and Emsley (2019).

Most studies find that both approaches mitigate motivational biases immediately after the training and retain their effects for several weeks (e.g., 8–12 weeks), with only a slight decline over time (e.g., Clegg et al., 2014; Morewedge et al., 2015; Symborski et al., 2014). In direct comparison, serious games generally prove more effective or at least as effective as educational videos with similar content (Morewedge et al., 2015; Symborski et al., 2014; Veinott et al., 2013).

While time is crucial in evaluating debiasing training effectiveness, an equally important question is whether participants transfer their knowledge to other business decisions. However, only two studies examine whether serious games mitigate motivational biases across different business contexts (Poos et al., 2017; Sellier et al., 2019a). Findings suggest that trained individuals successfully counter the confirmation bias but not the fundamental attribution error.[65] No research has explored this transfer effect for video trainings.

Beyond serious games, Poos et al. (2017) investigate whether the mere reading of a serious game script effectively mitigates biases. Their results show that participants mitigate the confirmation bias and fundamental attribution error as effectively as actual game players. Analogous to serious games, the confirmation bias effect transfers to unrelated tasks, but this does not hold for the fundamental attribution error. The authors attribute this finding to differences in task complexity: the fundamental attribution error transfer task encouraged superficial analysis, limiting transfer, while the confirmation bias task was shorter and simpler, making it easier to apply learned strategies (Poos et al., 2017).

Yoon et al. (2021) investigate the effectiveness of a training in which individuals observe others playing a serious game (i.e., observational training). While the investigated projection bias is reduced, the debiasing effect is weaker compared to playing the game (but still more pronounced compared to video trainings).

[65] Sellier et al. (2019a) investigate the confirmation bias through confirmatory hypothesis testing. Thus, the broad definition of confirmation bias is restricted to confirming hypothesis testing, raising the concern, whether the effects may result from the narrow operationalization of the confirmation bias.

Statistical Biases

The statistical biases examined in the training papers encompass representativeness bias, insensitivity to sample size, base rate neglect, stereotype bias, conjunction fallacy, gambler's fallacy, insensitivity to predictability, regression to the mean, and a combined statistical bias measured through a subset of these biases.

Whether serious games effectively mitigate statistical biases is tested using six biases. While five biases were reduced immediately after the training (i.e., representativeness bias, insensitivity to sample size, base rate neglect, stereotype bias, conjunction fallacy), gambler's fallacy was not. Further, only for the representativeness bias the mitigation effect was retained after several weeks.

Surprisingly, the training video approach is only applied to the representativeness bias. While it proves effective in both the short and long term, some studies suggest that serious game training is superior (Barton et al., 2016; Clegg et al., 2015). Similarly, observational training shows a significant reduction of the representativeness bias and is as effective as serious games or a simple training video (Yoon et al., 2021).

Studies on statistical biases have largely focused on awareness training. While eight out of nine biases have been examined, only four were consistently mitigated. Three studies find a reduction of the base rate neglect, four studies show a mitigation effect for the conjunction fallacy, and one study each finds a debiasing effect on the stereotype bias and insensitivity to predictability. While the training effect for the base rate neglect and conjunction fallacy is retained for at least eight weeks, this does not apply to the stereotype bias (Boissin et al., 2022). No significant mitigation effect is found for the regression to the mean, composite statistical bias, and insensitivity to sample size. For the gambler's fallacy, two studies report contradicting results—one finding a reduction, the other not (AlKhars et al., 2019; Lee et al. 2016). Comparing the awareness training to the serious games training, studies show that for the base rate neglect, and stereotype bias, both approaches are equally effective in bias mitigation.

Although analogical training (i.e., recognizing structural similarities across different situations) has been studied in only one study—and only for the composite statistical bias—the results show that the bias is still mitigated four weeks after the training (Aczel et al., 2015). Compared to awareness training, which shows no debiasing effect, analogical training seems to be a more effective intervention. The same study also examines whether participants transfer their debiasing knowledge on the composite statistical bias to real-life decisions (Aczel et al., 2015). Findings indicate that analogical training participants are more likely to report changed decision-making after four weeks than those in awareness training. However, as this effect is self-reported, further validation is needed.

Informational Biases

The informational biases addressed in the 13 studies reviewed are anchoring, ambiguity aversion, information bias, mental accounting error, overconfidence bias, overprecision bias, framing effect, outcome bias, and sunk cost fallacy.[66] While anchoring is well examined and considered by eleven studies, overconfidence is addressed in only two, and each of the remaining biases is investigated in just one study. Thus, while the finding that debiasing training reduces anchoring appears reliable, more research is needed for other biases. Based on the existing studies (and without considering the number of studies), mental accounting and overconfidence also appear to respond —at least in part— to training. Serious games effectively mitigate anchoring and mental accounting but have no impact on ambiguity aversion or information bias (Barton et al., 2016; Tommasi et al., 2021). Simpler training formats, such as videos (Clegg et al., 2015) and observational training (Yoon et al., 2021), also successfully reduce anchoring. Similar to the findings on motivational biases, serious games are at least as effective as video training. In addition, the debiasing effect of serious games and training videos on anchoring is retained several weeks after the intervention.

[66] Tommasi et al. (2021) also investigate the money illusion and zero risk bias. However, results on these two biases are not clearly reported in the paper and thus not discussed in this review.

The effectiveness of awareness training is mixed. For anchoring, two studies report bias reduction (Adame, 2016; Shepperd et al., 2018), while three others find no effect (Aczel et al., 2015; Lee et al., 2016; Welsh et al., 2007). Similarly, for overconfidence bias, one study finds a debiasing effect (Welsh et al., 2007), while another does not (Aczel et al., 2015). For all other informational biases tested, awareness training is ineffective (Aczel et al., 2015). Likewise, analogical training has no effect on any examined informational biases (Aczel et al., 2015).

Comparing not the type of training used, but the biases considered, one observation is sparking: The same type of training is not effective in mitigating all kinds of biases. While various trainings reduce motivational biases, statistical and informational biases are more challenging to mitigate. A potential explanation lies in their different drivers—statistical biases stem from misjudging correlations or the accuracy of outcomes, motivational biases arise from personal desires, and informational biases depend on information quality and quantity. These drivers involve distinct cognitive processes. We further explore factors influencing debiasing effectiveness in the mediating effects section.

[Insert Table 3 here]

Bias recognition and discrimination

The previous section focuses on the ultimate purpose of debiasing training, i.e., mitigating a bias. However, there are situations where it is equally important to recognize and correctly classify a bias. For example, management accountants are responsible for ensuring the rationality of decision-making processes by managers. If decision-makers have not undergone debiasing training, the management accountant must be able to identify a potential bias and offer help (to overcome the bias). Another example are decision-makers relying on third-party advice (e.g., during M&A processes). Decision quality is higher if decision-makers recognize that the advice is subject to cognitive biases. A similar case could be made for financial advisors recognizing that their client has fallen prey to a bias.

Eleven studies in this review examine bias recognition (i.e., knowledge) and discrimination (i.e., differentiating between biases) (see Table 4). These studies typically employ questionnaires to assess bias recognition. Participants are asked to identify a bias in a hypothetical scenario and select the relevant bias from the list of biases that were previously covered in the training (Morewedge et al., 2015; Yoon et al., 2021). The studies in this field consider six different biases (i.e., confirmation bias, fundamental attribution error, blind spot bias, projection bias, representativeness bias, and anchoring). Participants' recognition and discrimination ability is increased after playing a serious game or watching a training video. In line with our findings in the previous section, the effect of training videos and serious games is retained for several weeks. Surprisingly, however, the training video outperforms (or is at least as effective as) serious games.[67]

One study examines observational training (Yoon et al., 2021). This study finds a significant improvement in bias recognition and discrimination for all three biases examined in that study (i.e., projection bias, representativeness bias, and anchoring effect). The effect is also retained for several weeks. While the observational training is equally effective as serious games, the training video is again more effective.

In conclusion, the results show that debiasing training effectively increases bias awareness and knowledge. Notably, video training is more effective in increasing bias knowledge, while serious games are superior in mitigating biases. This difference stems from their learning approaches: videos provide a concentrated, static dose of information, whereas serious games engage participants interactively, allowing them to learn through mistakes over time. According to constructivist learning theory, memorization alone hinders the formation of meaningful associations and the cognitive processes essential for problem-solving (Rooney,

[67] Contradicting results are found by Symborski et al. (2014), who report that the serious game is superior to the video training, and by Clegg et al. (2014) and Clegg et al. (2015), who find that a repeated game play and repeated and single game play taken together outperform the video training.

2012; Symborski et al., 2014). Thus, a key insight is that bias knowledge does not equally translate into bias mitigation, and different training methods serve distinct purposes. We further explore the practical implications of these findings in the discussion section.

[Insert Table 4 here]

Moderating effects

While numerous studies show that serious games effectively mitigate biases, ten (out of 19) also explore important moderators, and one study examines moderators in awareness trainings. In what follows we discuss these moderating effects, summarized in Table 5 (serious games) and Table 6 (awareness trainings). We first discuss the six moderators explored by prior research on serious games, i.e., *duration, repetition, player type, feedback timing, game style,* and *rewards*.

Duration

It could be assumed that playing a serious game for a longer period of time improves bias mitigation. Potentially, as players become more familiar with the game, the cognitive load associated with learning how to operate the game decreases and allows them to focus more on the training content (Dunbar et al., 2014a). Additionally, extended exposure may strengthen retention of the learning material.

Yet, three out of four studies find no effect for the biases investigated (i.e., confirmation bias, fundamental attribution error, and blind spot bias) (Bessarabova et al., 2016; Dunbar et al., 2017; Veinott et al., 2013).[68] A potential explanation is that players experience mental fatigue and suffer from cognitive overload (Gonzalez et al., 2011). This could lead to a reliance on heuristic processing, which may offset any additional positive effects gained from prolonged exposure (Dunbar et al., 2014a). Thus, serious games are effective even when they last only for 30 minutes and their effectiveness mainly does not increase with longer gameplay.

[68] Only Dunbar et al. (2014a) find an advantage of a longer game duration (i.e., 60 minutes instead of only 30 minutes) for the confirmation bias and fundamental attribution error.

Repetition

Seven studies investigate whether repeated gameplay (instead of longer duration) improves bias mitigation by reinforcing knowledge. However, results are again mixed. Repeated gameplay positively affects confirmation bias (three of five studies) and anchoring (one study) (Clegg et al., 2014; Clegg et al., 2015; Dunbar et al., 2017; Dunbar et al., 2014a) but does not improve performance on the fundamental attribution error or blind spot bias (Bessarabova et al., 2016; Clegg et al., 2014; Dunbar et al., 2017; Dunbar et al., 2014b; Veinott et al., 2013). Findings on projection and representativeness bias are less clear—playing the game once yields to better immediate effects, while repeated gameplay improves retention over several weeks (Clegg et al., 2015).

One might argue that repeated exposure is more advantageous for anchoring and confirmation bias because both biases have a straightforward mechanism, where being repeatedly prompted to question initial assumptions (anchoring) or check for counter-evidence (confirmation bias) reinforces debiasing. In contrast, fundamental attribution error and blind spot bias are more complex and deeply rooted in self-perception and social cognition, making them more challenging to mitigate further through repeated gameplay alone.

Player type

Another moderator considered is the number of players, i.e., multi-player versus single-player mode (Bessarabova et al., 2016; Dunbar et al., 2017; Dunbar et al., 2014a). In the single-player mode, participants play the serious game individually, whereas attendants of the multi-player game work in teams of two, advise each other, and must agree on a joint answer (Dunbar et al., 2014a). While a multiplayer mode is predicted to be superior in mitigating the confirmation bias and fundamental attribution error (Dunbar et al., 2017; Dunbar et al., 2014a), the contrary is assumed for the blind spot bias (Bessarabova et al., 2016). Concerning the blind spot bias, talking about and admitting one's own biases in a multi-player mode may lead to ego defense and self-enhancement, activating the process of over-relying on one's introspective information, i.e., the exact process that is intended to combat (Bessarabova et al., 2016). For the confirmation bias and fundamental attribution error, individuals profit from a higher learning level by explaining their reasoning and joint decision-making (Bessarabova et al., 2016).

However, these predictions are not supported. Interestingly, a single-player mode is superior to a multi-player mode in mitigating the confirmation bias (Dunbar et al., 2017; Dunbar et al., 2014a). A potential explanation is that while communication and collaboration may activate System 2 thinking, it also distracts from learning (Dunbar et al., 2017).[69]

Feedback timing

Two studies investigate how feedback timing, i.e., immediate feedback after a response or accumulated after a set of questions, affects bias mitigation (Bessarabova et al., 2016; Dunbar et al., 2017). It is argued that immediate feedback might result in faster learning and error correction, leading to improved performance. However, both studies find no difference for the biases investigated (i.e., confirmation bias, fundamental attribution error, and blind spot bias). While immediate feedback ensures learning, it also interrupts the gameplay and slows down the immersive experience (Bessarabova et al., 2016). The latter (negative) effect likely offsets the positive effect of increased learning through immediate feedback.

Game style

The largest number of moderators falls into the category of (serious) game style, which includes studies on confirmation bias, fundamental attribution error, and blind spot bias. Researchers have examined whether specific game design elements improve debiasing, including (a) *designing a personalized avatar* by customizing gender and body shape (Shaw et al., 2018), (b) incorporating a more *detailed art style* with rich and realistic visuals (Martey et al.,

[69] Related thereto, one study finds that feedback directly provided by the game (in a pop-up window) is more effective in mitigating the confirmation bias than feedback from team members in the multi-player version (Dunbar et al., 2017). Similarly, this may be because multi-player interactions create distractions, reducing players' focus on the debiasing process (Dunbar et al., 2017).

2017), (d) shifting from a first-person to a *third-person camera perspective,* placing the player behind the avatar rather than inside it (Veinott et al., 2013), and (e) enhancing the game with *explicit training* elements, such as introduction to cognitive biases, bias definitions, and quizzes) (Bessarabova et al., 2016; Dunbar et al., 2014b). These game style moderators assume to increase players' involvement (a), avatar credibility (b), engagement and self-reflection (c), reorientation towards the environment (d), and reduce cognitive load (e). In turn, this is expected to enhance exploratory learning experiences and knowledge transfer, ultimately improving bias mitigation. However, among these moderators, only (e) explicit training effectively enhances bias mitigation, and even then, only for the confirmation bias (Bessarabova et al., 2016; Dunbar et al., 2014b). This may be due to the complex nature of confirmation bias, which relies on belief-confirmation patterns that benefit more directly from increased knowledge and awareness, making explicit instruction particularly effective. Interestingly, contrary to initial predictions, a (b) minimalistic art style is slightly more effective for mitigating blind spot bias, and a (c) minimal narrative style is slightly superior for reducing the confirmation bias (Martey et al., 2017). These results suggest that excessive detail and rich narratives may introduce distractions rather than enhance learning.

Rewards

The final moderator for serious games is the number of *rewards* granted during gameplay. One study examines whether intangible rewards—such as badges, victory sounds, virtual prizes, and visual feedback—enhance bias avoidance (McKernan et al., 2015). High rewards are thought to foster accomplishment, encouraging reflection and motivation. Reward-based feedback enhances players' sense of competence and self-determination, increasing engagement (McKernan et al., 2015). Thus, more rewards are expected to improve learning. However, results show no effect for the number of rewards on mitigating the projection bias, anchoring, and representativeness bias (McKernan et al., 2015). The study finds that the feeling of being rewarded is not related to the quantity of rewards. Instead, players may already feel rewarded through game progression and successful challenges, making additional rewards less impactful. Consequently, learning outcomes remain unaffected by reward quantity.

[Insert Table 5 here]

Beyond serious games, one study examines moderators for awareness training, i.e., *decision time* and *repetition* on base rate neglect and the conjunction fallacy. Within the moderator category *decision time*, the study examines whether a short single-shot explanation of biases and a debiasing strategy improve bias mitigation when participants have short vs. long response times. Participants first provide an immediate, intuitive response to a cognitively demanding question measuring a bias and then answer the same question deliberately with no time constraint. Surprisingly, most participants answered correctly right away when given a short decision time, eliminating the need for correction during the long decision phase (Boissin et al., 2022). This suggests that a short decision time after the training intervention is sufficient to mitigate biases. Once biases are properly explained, individuals respond rationally, even under time pressure. The same study also examines whether *repetition* leads to greater debiasing effects. It is argued that a second training increases learning and fosters debiasing. However, the results do not support this prediction (Boissin et al., 2022).

[Insert Table 6 here]

Mediating effects

To better understand why some training interventions are effective while others are not, we examine mediating effects. However, the reviewed studies barely investigate process variables and often lack theoretical reasoning. Therefore, we propose theories aligned with the study results and identify potential mediators explaining the presence or absence of debiasing effectiveness. These mediators are classified as *training variables* or *personal variables*. Training variables influence debiasing through mechanisms initiated by the training, while personal variables are effective via a change in perception. Figure 1 shows the mediators and moderators.

Training variables

Most studies investigate serious games and often compare them to video training. Overall, serious games seem superior (Clegg et al., 2014; Clegg et al., 2015; Dunbar et al., 2017; Dunbar et al., 2014a; Morewedge et al., 2015; Rhodes et al., 2017; Shaw et al., 2018; Veinott et al., 2013). A first explanation for this is greater *interaction* and *involvement*—players must actively make decisions that directly impact the game, fostering responsibility and involvement (Bell & Kozlowski, 2008; Frese et al., 1991; Sitzmann, 2011). Thus, players are more (emotionally) engaged in the learning experience (Garris et al., 2002; O'Neil & Chen, 2005; Sitzmann, 2011). Yoon et al. (2021) further support this, showing that participants are more engaged in serious games than in training videos and observational training.[70]

Theorists argue that engagement enhances learning by promoting metacognition and encouraging greater cognitive effort in evaluating, memorizing, and integrating new information (Bell & Kozlowski, 2008; Sitzmann et al., 2006). Active learning, defined as engaging individuals in "[…] doing things and thinking about what they are doing" (Bonwell & Eison, 1991, p. 5), supports deep learning (Peters, 2011). It helps players develop a mental model of the training content, improving their ability to apply it to other contexts (Bell & Kozlowski, 2008; Healy & McCutcheon, 2008; Keith & Frese, 2005). This also aligns with Shaw et al. (2018), who interviewed serious game participants. One interviewee stated: "[…] playing the game, you are a lot more active, and if you're listening to a lecture, reading a book, you can kinda get away with not having to learn it. Whereas the game – like it makes you learn it, and makes you understand it. I think that was a lot more beneficial." (Shaw et al., 2018, p. 23).

Thus, active learning—as in serious games—better engages trainees, which enhances

[70] However, study results also show that for two out of three biases observational training is equally effective to serious games (Yoon et al., 2021). By questioning whether the observed person's decisions are biased, observers actively reflect on different biases and develop a greater interest in debiasing techniques for their own decisions. This might compensate for the lack of interaction to deepen the learned concepts and the decreased excitement of observing rather than playing a game.

learning. In contrast, passive trainings like video training, where participants simply receive information, lack this advantage (Garris et al., 2002; Riley & Ward, 2017; Yoon et al., 2021).[71] *Failure allowance* is a second mediator that might explain the superiority of serious games. According to experiential learning theory (Kolb, 2015)—also known as the "trial and error" approach (Gentry, 1990)—individuals learn through experience, reflection, thinking, and action (Kolb, 2015). Players first encounter various scenarios and make decisions, followed by feedback and reflecting on the outcomes. After a mistake, players are encouraged to think critically and form conclusions, which they can verify during the next action phase. A participant in the study of Shaw et al. (2018) stated after playing a serious game "[Y]ou had to figure out – and I wasn't just gonna give up and not figure out what I had to do." (Shaw et al., 2018, p. 23). Serious games are by nature interactive and allow for safe failure, providing participants with a rich learning experience that deepens understanding and retention.

Next, *simplification* can be understood as another mediator. The concept and underlying principles of cognitive biases are abstract and difficult to access. Based on cognitive load theory, simplifying abstract bias knowledge into easily understandable context allows learners to process and retain information more effectively through which they are more likely to recognize and battle biases, especially in the long run (Sweller, 1994).

Serious games, for instance, draw on entertaining crime scenarios in which the player has to identify a murderer. By describing cognitive biases and how to mitigate them using easy and real-world examples, instead of definitions and conceptual drawings, participants are more likely to internalize and transfer the knowledge. This simplification also supports the effectiveness of observational training (Yoon et al., 2021) and the training by reading the game script, which shows to be as effective as actually playing the game (Poos et al., 2017). While

[71] For awareness training, it is less clear that this training is passive. While the core of awareness training is always the provision of text-based explanations, they may also integrate active parts such as practice sessions and feedback rounds (Adame, 2016; Cassotti & Moutier, 2010; van Brussel et al., 2021).

we emphasize the benefits of interactivity and the opportunity for safe failure in enhancing learning, it appears that merely simplifying the abstract principles of biases and debiasing strategies within a game format is also effective for bias mitigation. Similarly, findings from awareness training suggest that breaking down the abstract concept into simple explanations—such as providing an example question along with both biased and unbiased responses—facilitates understanding and thus aids in reducing biases (Boissin et al., 2022; Lee et al., 2016).

Personal variables

A key personal-level mediator across all training interventions is *awareness*. Effective debiasing training must raise bias awareness and the benefits of debiasing (Fischhoff, 1982), as increased awareness reduces biased decision-making. Aligned with constructivism theory—which suggests that individuals construct knowledge and skills through experience and reflection—awareness enables learners to contextualize their knowledge and develop a deeper understanding (Bada & Olusegun, 2015; Rooney, 2012; Symborski et al., 2014). Only by recognizing how biases influence learners' choices and the consequences they entail can individuals effectively mitigate them in decision-making. Since all training approaches enhance awareness to some extent, this may partly explain their effectiveness in reducing biases.

Self-determination may also explain the superiority of serious games over other training interventions. According to the andragogical principle of self-concept, individuals learn best through self-directed learning (Knowles, 1990; Symborski et al., 2014). Serious games support this by allowing players to progress at their own pace, revisit content, and explore different outcomes. This flexibility prevents cognitive overload by delivering information in stages (Sweller, 1994) and enhances motivation and knowledge retention (Clegg et al., 2015; Snowden & Halsall, 2016). By giving learners control over their learning experience, serious games foster deeper learning and more effective training outcomes compared to passive video-based approach where content is presented only once, requiring immediate comprehension.[72]

Lastly, research has shown that intrinsic motivation and, thus, task *enjoyment* significantly impact learning outcomes (Garris et al., 2002; Sitzmann, 2011). Enjoyment fosters engagement, encourages deeper thinking to build new knowledge (Hernik & Jaworska, 2018) and creates a "flow state"—an optimal performance condition in which engagement is at its peak (Csikszentmihalyi, 2009; Sitzmann, 2011). Serious games are inherently motivating, offering enjoyable learning experiences through challenges, curiosity, and fantasy (Malone, 1981; Sitzmann, 2011). Studies have shown that embedding learning content within a fantasy context increases interest and learning more than traditional training methods (Garris et al., 2002; Sitzmann, 2011; Wilson et al., 2009). Interviews confirm this preference, with participants reporting greater enjoyment and learning from games than from video training (Shaw et al., 2018).

Further, challenges in serious games spark curiosity and intrinsic motivation (Sitzmann, 2011), enhancing learning and improving debiasing. Enjoyment further sustains engagement, increasing participants' willingness to invest time in continued training (Sitzmann, 2011). This may be essential for retaining debiasing effects beyond the 8-12 weeks examined in studies.

[Insert Figure 1 here]

IV. Discussion

Our review provides an overview of several training formats. From a practitioner's perspective, this may help firms and particularly management accountants to determine which training design best fits their needs. Among other aspects, this choice depends on the biases most important for the firm. However, since studies focus on selected biases rather than all, the picture remains incomplete. Thus, a recommendation for a one-size-fits-all training approach

[72] While one could argue that training videos can also be paused and replayed, studies do not indicate whether participants were able to do so. In contrast, studies report video duration as an exact value rather than a mean (e.g., Clegg et al., 2014), explicitly state that the video was watched only once (e.g., Dunbar et al., 2017), or state that all participants viewed it at the same rate (e.g., Shaw et al., 2018). Hence, we conclude that training videos lack self-determination.

is not feasible. Below, we discuss important takeaways and directions for future research.

Training Effectiveness and Efficiency

Research on debiasing training largely focuses on serious games, which effectively mitigate eleven biases (of 14 biases examined) and maintain their impact for several weeks. Yet, from an efficiency perspective, the benefits must be weighed against costs. License-based serious games, especially those tailored to firm-specific biases, can be costly and require significant time investment for setup, explanation, and gameplay (CTI, 2025). For instance, the duration of serious games can range from 30 minutes (e.g., Bessarabova et al., 2016; Dunbar et al., 2014b; Shaw et al., 2018) to up to 4 hours (e.g., Poos et al., 2017).

Video training is a resource-efficient alternative, requiring less time for instruction, programming and conduction. This may explain why firms like Google or Meta use it for debiasing. After serious games, training videos are the most researched and appear reliable—at least for motivational biases, as research on other biases are scarce. Compared to serious games, training videos are less effective in mitigating biases. A possible explanation is that active decision-making with feedback enhances learning more than passively absorbing information.

Observational training, where individuals watch others play a serious game, offers a similar cost-benefit ratio to direct gameplay. While achieving comparable debiasing effects, observational training has the advantage that two employees can be trained simultaneously with one game license. Script-based training, where participants read a serious game script, also provides comparable debiasing effects to gameplay. Since script-based training requires no specialized IT equipment or licensing fees, it may be a cost-efficient choice for practitioners. However, the limited number of studies and biases tested by these two game-related approaches make clear recommendation difficult. Additionally, both approaches lack interactivity and enjoyment, which may reduce employees' willingness to participate or repeat the training.

Awareness trainings are another cost-efficient alternative that effectively mitigate (some) statistical biases but has little impact on motivational and informational biases. Despite its limitations, firms may still favor it due to ease of implementation and low resource requirements. One possible explanation why studies find no mitigating or mixed effects for some biases is that this simple training intervention becomes less effective when the required debiasing strategies, e.g., using Bayesian reasoning, become more complex and abstract (Larrick, 2004). The lack of practice opportunities and feedback further limits learning, and low interaction may reduce motivation. Given the ineffectiveness of awareness training for informational biases, these challenges seem more important for mitigating informational than statistical biases. A possible solution is suggested by Lee et al. (2016) who find that combining a slide show with a serious game yields the strongest mitigation effects. Future research should apply this simple form of debiasing training in combination with more interactive training approaches to see whether all three categories of biases can be more effectively mitigated.

Analogical training, which teaches biases through structural similarities across different situations, has shown limited effectiveness. It mitigates the composite statistical bias but none of the informational biases examined (Aczel et al., 2015).[73] On the one hand, the strength of this intervention lies in training abstract principles such as statistical rules (Aczel et al., 2015). On the other hand, however, as informational biases are less abstract, analogical training is less effective for these. For instance, statistical biases require probabilistic thinking and abstract reasoning (e.g., Bayesian reasoning), whereas informational biases stem from misinterpreting or misusing data. An example is the framing effect, where deviations from rational decision-making arise due to differences in wording and perception rather than statistical reasoning. Nonetheless, a word of caution is needed for appliers. First, analogical training requires group discussions, preventing self-learning and increasing complexity. Second, as only one study has examined its debiasing effects, the generalizability is limited.

[73] Motivational biases have not been examined.

Moderators and Mediators

Besides analyzing the effectiveness of various training approaches, we provide an overview of moderators and (potential) mediators. The majority of moderators have been examined for serious games. The focus of this research is whether more-complex design elements of serious games (e.g., multi-player mode, narrative richness of the game, or avatar choice) increase effectiveness. Interestingly, however, simpler forms generally prove more effective. The only more complex design elements with mixed findings, i.e., where it is not clear whether a simple design is better, are explicit training, game duration and repetition. At least for some biases enhancing gameplay with additional definitions and quizzes (Dunbar et al., 2014b), playing the game for a longer duration (i.e., 60 minutes) (Dunbar et al., 2014a) and playing it twice instead of once (Clegg et al., 2014; Clegg et al., 2015; Dunbar et al., 2017; Dunbar et al., 2014a) increases bias mitigation. These findings suggest that the debiasing effects are mainly driven by the content and structure of the game itself, i.e., how it presents challenges and scenarios that trigger reflection, rather than by complex design features.

Contrary to serious games, the (single) study on moderating effects in awareness training finds that repetition does not enhance debiasing (Boissin et al., 2022). However, it shows that both intuitive and deliberate decisions become more rational after training. This suggests that simply reading definitions and mitigation techniques can significantly improve both fast, intuitive reasoning (System 1) and slow, deliberate thinking (System 2) (Boissin et al., 2022).

While most moderators appear to have little impact, practitioners should focus on the mediating factors that drive the debiasing effect. On a *training level*, interaction and engagement, the allowance for failure, the simplification of abstract learning content and the ability to independently detect and recognize biases drive the effectiveness of trainings. On a *personal level*, fostering bias awareness, allowing self-determination and creating enjoyable experiences likely deepens the knowledge gained in a training.

Mediators are particularly prominent in serious games, contributing to their superiority over other training interventions (Yoon et al., 2021). Serious games interactive nature fosters high engagement and enjoyment, allowing individuals to uncover biases independently rather than passively receiving information. This enhances learning retention and transferability to other bias-prone situations. In contrast, video or awareness training lack interaction, enjoyment, allowance for failure, and self-determination. This can be disadvantageous when employees fail to recognize the value of debiasing and are hesitant to invest time.

Training Recommendations

The analysis above results in some recommendations for practitioners. First, a firm must evaluate whether bias recognition or mitigation should be the primary focus of training. When firms heavily rely on third-party advice (e.g., advisors for determining transfer prices) or provide advisory services themselves (e.g., financial advisors), bias recognition is crucial. In such cases, training videos should be prioritized over serious games, as they have proven to be more effective for bias recognition. However, when the goal is to mitigate decision-makers own biases, management accountants should offer help by introducing debasing strategies, such as training interventions. Across all bias categories, serious games have shown the most consistent results in bias mitigation and should therefore be the preferred training approach.

Beyond these recommendations, the complexity of training and employees' (intrinsic) motivation to participate must be considered. For highly motivated employees, simpler, cost-effective methods like video, awareness, or observational training may suffice. Otherwise, interactive approaches such as serious games and related interventions are recommended. Since debiasing is a new concept for many employees, we suggest using serious games as an initial training, followed by simpler methods to reinforce and retain the content after several weeks.[74]

Second, firms should assess whether new knowledge transfers to daily business decisions

[74] This is in line with Lee et al. (2016) who show that the combination of a serious game and simpler trainings is more effective than each training individually.

(Baldwin & Ford, 1988; Sitzmann, 2011). Despite its importance, only three studies examine whether trained participants apply their knowledge to untrained situations, and all face measurement issues (Aczel et al., 2015; Poos et al., 2017; Sellier et al., 2019b). Aczel et al. (2015) show that, participants self-report changes in decision-making after analogical (58%) or awareness trainings (32.43%) for one bias.[75] Poos et al. (2017) find that training effects (via serious games or reading the script) on bias mitigation transfer for one bias but not for the other. While the authors claim to use questions unrelated to the game and script, these questions still appear similar to the ones used in the training context. Sellier et al. (2019b) report that 19% of participants trained with a serious game successfully transfer their knowledge to a different problem and context in a field setting.

Third, the frequency of training interventions must be carefully determined. Study results show that, to some extent, repeating training enhances debiasing effects and promotes long-term learning outcomes, but findings are mostly limited to two- to three-month periods. However, frequent training may reduce acceptance and, thus, increased employee resistance.

Avenues for Future Research

Based on the insights drawn from the review and the discussion above, fruitful avenues for future research can be identified. First, more research is needed to identify the conditions under which debiasing knowledge is retained and can be transferred to untrained situations. In prior studies, long-term effectiveness is often limited to a couple of weeks. Moreover, transfer effects are examined by just a few studies, and their findings should be interpreted with caution due to measurement limitations, such as reliance on subjective assessments (Aczel et al., 2015; Poos et al., 2017; Sellier et al., 2019a). From a theory perspective, it is still unclear what determines whether debiasing knowledge can be transferred to untrained contexts.

[75] However, the measurement of this effect is based on a single self-reported question, asking participants whether they decided differently in real-life situations within the four weeks after the training.

Second, the number of biases examined in prior training papers is limited. Future research should examine whether other biases which have either not been investigated (e.g. status-quo bias) or could not be mitigated (e.g., sunk cost fallacy) are reduced by a new or modified training approach. Existing research does not provide a compelling answer to the question why a specific training is effective in mitigating (or recognizing) one bias but not the other. Research in this field could help to further improve debiasing training.

Third, while some research explores moderators for serious games, there is still a lack of research on moderators (and mediators) for other training interventions. The mediating effects suggested in this review are theory-based but remain untested empirically. Thus, further research is needed to identify which training features drive debiasing effects.

Fourth, the interaction of debiasing training and other management controls is largely unexplored. For example, research could examine whether combining incentives with training improves effectiveness over training alone. This may occur if debiasing knowledge fails to transfer to untrained situations due to insufficient (extrinsic) motivation.

V. Conclusion

Cognitive biases are prevalent in corporate decision-making and can have detrimental effects on firm performance (Soll et al., 2015). The role of management accountants is to identify opportunities for debiasing and systematically enhance the quality of management decisions. One approach is debiasing training. This review examines research in the field, discusses key findings, and identifies avenues for future research. Our findings indicate that while various types of training exist, there is no first-best approach. Rather, the type of training should align with the firm’s specific needs, its’ budget constraints, and the nature of relevant biases.

Our review covers research on six training approaches. While some overlap exists, research on the effectiveness of these approaches often focus on different biases, making direct comparisons challenging. Additionally, studies vary in their dependent variable, assessing bias recognition (which might be more important for management accountants or advisors) and/or bias mitigation (which matters for all decision-makers).

In terms of bias mitigation, serious games often prove favorable. This appears to be rooted in the interactive and immersive nature of serious games, which engages participants in problem-solving and decision-making. While highly motivating, they also come with high implementation costs (Sitzmann, 2011). More cost-efficient yet—compared to serious games—generally less effective alternatives include video and awareness trainings, with the latter showing particularly high variability in bias mitigation. Remarkably, video trainings outperform serious games in bias recognition but not mitigation. A cost-benefit compromise for firms may be game-related trainings like reading serious game scripts or observational training.

From a theory perspective, the superiority of serious games over more passive formats such as video training aligns with cognitive load theory, indicating that trainings that actively involve participants help retaining and transferring this knowledge to new contexts. This provides a foundation for developing training interventions that balance cognitive demand with meaningful engagement, especially for complex cognitive biases. The mediators identified in the review (e.g., failure allowance, interaction and enjoyment) have practical implications, suggesting that integrating these elements into training formats may improve mitigation.

Our review also offers directions for future research. In a nutshell, we make different suggestion for future research that aim at a better understanding of what determines the effectiveness of a specific training format.

References

Aczel, B., Bago, B., Szollosi, A., Foldes, A., & Lukacs, B. (2015). Is it time for studying real-life debiasing? Evaluation of the effectiveness of an analogical intervention technique. *Frontiers in Psychology*, *6*, Article 1120, 1–13. https://doi.org/10.3389/fpsyg.2015.01120

Adame, B. J. (2016). Training in the mitigation of anchoring bias: A test of the consider-the-opposite strategy. *Learning and Motivation*, *53*, 36–48. https://doi.org/10.1016/j.lmot.2015.11.002

AlKhars, M., Evangelopoulos, N., Pavur, R., & Kulkarni, S. (2019). Cognitive biases resulting from the representativeness heuristic in operations management: An experimental investigation. *Psychology Research and Behavior Management*, *12*, 263–276. https://doi.org/10.2147/PRBM.S193092

Arkes, H. R. (1991). Costs and Benefits of Judgment Errors: Implications for Debiasing. *Psychological Bulletin*, *110*(3), 486–498. https://doi.org/10.1037/0033-2909.110.3.486

Arkes, H. R., & Blumer, C. (1985). The Psychology of Sunk Cost. *Organizational Behavior and Human Decision Processes*, *35*(1), 124–140. https://doi.org/10.1016/0749-5978(85)90049-4

Bada, D., & Olusegun, S. (2015). Constructivism learning theory: A paradigm for teaching and learning. *Journal of Research & Method in Education*, *5*(6), 66–70.

Baer, T., Heiligtag, S., & Samandari, H. (2017). *The business logic in debiasing*. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/the-business-logic-in-debiasing

Baldwin, T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, *41*(1), 63–105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Baron, J., Beattie, J., & Hershey, J. C [John C.] (1988). Heuristics and biases in diagnostic reasoning. *Organizational Behavior and Human Decision Processes*, *42*(1), 88–110. https://doi.org/10.1016/0749-5978(88)90021-0

Baron, J., & Hershey, J. C [J. C.] (1988). Outcome Bias in Decision Evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569–579. https://doi.org/10.1037//0022-3514.54.4.569

Barton, M., Symborski, C., Quinn, M., Morewedge, C. K., Kassam, K. S., & Korris, J. H. (2016). The Use of Theory in Designing a Serious Game for the Reduction of Cognitive Biases. *Transactions of the Digital Games Research Association (ToDIGRA)*, *2*(3), 61–87. https://doi.org/10.26503/todigra.v2i3.53

Beaulac, G., & Kenyon, T. (2018). The Scope of Debiasing in the Classroom. *Topoi*, *37*(1), 93–102. https://doi.org/10.1007/s11245-016-9398-8

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *The Journal of Applied Psychology*, *93*(2), 296–316. https://doi.org/10.1037/0021-9010.93.2.296

Berthet, V. (2021a). The Impact of Cognitive Biases on Professionals' Decision-Making: A Review of Four Occupational Areas. *Frontiers in Psychology*, *12*, 802439. https://doi.org/10.3389/fpsyg.2021.802439

Berthet, V. (2021b). The Measurement of Individual Differences in Cognitive Biases: A Review and Improvement. *Frontiers in Psychology*, *12*, 630177. https://doi.org/10.3389/fpsyg.2021.630177

Bessarabova, E., Piercy, C. W., King, S., Vincent, C., Dunbar, N. E., Burgoon, J. K., Miller, C. H., Jensen, M., Elkins, A., Wilson, D. W., Wilson, S. N., & Lee, Y.-H. (2016). Mitigating bias blind spot via a serious video game. *Computers in Human Behavior*, *62*, 452–466. https://doi.org/10.1016/j.chb.2016.03.089

Boissin, E., Caparos, S., Voudouri, A., & Neys, W. de (2022). Debiasing System 1: Training favours logical over stereotypical intuiting. *Judgment and Decision Making*, *17*(4), 646–690. https://doi.org/10.1017/S1930297500008895

Bonwell, C. C., & Eison, J. A. (1991). Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports. *0884-0040*. https://eric.ed.gov/?id=ED336049

Brüggen, A., & Luft, J. L. (2016). Cost Estimates, Cost Overruns, and Project Continuation Decisions. *The Accounting Review*, *91*(3), 793–810. https://doi.org/10.2308/accr-51202

Bush, R. M. (2017). Serious Play: An Introduction to the Sirius Research Program. *Games and Culture*, *12*(3), 227–232. https://doi.org/10.1177/1555412016675728

Cassotti, M., & Moutier, S. (2010). How to explain receptivity to conjunction-fallacy inhibition training: Evidence from the Iowa gambling task. *Brain and Cognition*, *72*(3), 378–384. https://doi.org/10.1016/j.bandc.2009.11.004

Chang, W., Chen, E., Mellers, B., & Tetlock, P. (2016). Developing expert political judgment: The impact of training and practice on judgmental accuracy in geopolitical forecasting tournaments. *Judgment and Decision Making*, *11*(5), 509–526. https://doi.org/10.1017/S1930297500004599

Clegg, B. A., Martey, R. M [Rosa M.], Stromer-Galley, J., Kenski, K., & Saulnier, T. (2014). Game-based Training to Mitigate Three Forms of Cognitive Bias. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Article 14180, 1–12.

Clegg, B. A., McKernan, B., Martey, R. M [Rosa M.], Taylor, S. M., Stromer-Galley, J., Kenski, K., Saulnier, E. T [E. Tobi], Rhodes, M. G., Folkestad, J. E., McLaren, E., Shaw, A., & Strzalkowski, T. (2015). Effective Mitigation of Anchoring Bias, Projection Bias, and Representativeness Bias from Serious Game-based Training. *Procedia Manufacturing*, *3*, 1558–1565. https://doi.org/10.1016/j.promfg.2015.07.438

Csikszentmihalyi, M. (2009). *Flow: The psychology of optimal experience* [Nachdr.]. *Harper Perennial Modern Classics*. Harper [and] Row.

CTI. (2025). *Creative Technologies, Inc. – Immersive simulation for government and industry*. https://www.cretecinc.com/

Ding, S., & Beaulieu, P. (2011). The Role of Financial Incentives in Balanced Scorecard-Based Performance Evaluations: Correcting Mood Congruency Biases. *Journal of Accounting Research*, *49*(5), 1223–1247. https://doi.org/10.1111/j.1475-679X.2011.00421.x

Dunbar, N. E., Jensen, M. L., Miller, C. H., Bessarabova, E., Lee, Y.-H., Wilson, S. N., Elizondo, J., Adame, B. J., Valacich, J., Straub, S., Burgoon, J. K., Lane, B., Piercy, C. W., Wilson, D., King, S., Vincent, C., & Schuetzler, R. M. (2017). Mitigation of Cognitive Bias with a Serious Game: Two Experiments Testing Feedback Timing and Source. *International Journal of Game-Based Learning*, *7*(4), 86–100. https://doi.org/10.4018/IJGBL.2017100105

Dunbar, N. E., Jensen, M. L., Miller, C. H., Bessarabova, E., Straub, S. K., Wilson, S. N., Elizondo, J., Burgoon, J. K., Valacich, J. S., Adame, B., Lee, Y.-H., Lane, B., Piercy, C., Wilson, D., King, S., Vincent, C., & Scheutzler, R. (2014a). Mitigating Cognitive Bias through the Use of Serious Games: Effects of Feedback. *Persuasive Technology - 9th International Conference*, *8462*, 92–105. https://doi.org/10.1007/978-3-319-07127-5_9

Dunbar, N. E., Miller, C. H., Adame, B. J., Elizondo, J., Wilson, S. N., Lane, B. L., Kauffman, A. A., Bessarabova, E., Jensen, M. L., Straub, S. K., Lee, Y.-H., Burgoon, J. K., Valacich, J. J., Jenkins, J., & Zhang, J. (2014b). Implicit and explicit training in the mitigation of cognitive bias through the use of a serious game. *Computers in Human Behavior*, *37*, 307–318. https://doi.org/10.1016/j.chb.2014.04.053

Ehrlinger, J., Readinger, W. O., & Kim, B. (2016). Decision-Making and Cognitive Biases, 5–12. https://doi.org/10.1016/B978-0-12-397045-9.00206-8

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgement under Uncertainty: Heuristics and Biases* (pp. 422–444). Cambridge University Press.

Fox, C. R., & Tversky, A [A.] (1995). Ambiguity Aversion and Comparative Ignorance. *The Quarterly Journal of Economics*, *110*(3), 585–603. https://doi.org/10.2307/2946693

Frese, M., Brodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in Training Computer Skills: on the Positive Function of Errors. *Human–Computer Interaction*, *6*(1), 77–93. https://doi.org/10.1207/s15327051hci0601_3

Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, *33*(4), 441–467. https://doi.org/10.1177/1046878102238607

Gentry, J. W. (1990). What is experiential learning? In J. W. Gentry (Ed.), *Guide to business gaining and experiential learning* (pp. 9–20). Kogan Page.

Gilbert, D. T. (1998). Ordinary Personology. In D. T. Gilbert, T. Fiske, & G. Lindzey (Eds.), *The Handbook of Social Psychology* (4th ed., pp. 89–150). McGraw-Hill.

Gonzalez, C., Best, B., Healy, A. F., Kole, J. A., & Bourne, L. E. (2011). A cognitive modeling account of simultaneous learning and fatigue effects. *Cognitive Systems Research*, *12*(1), 19–32. https://doi.org/10.1016/j.cogsys.2010.06.004

Google. (2013). *Raise awareness about unconscious bias*. https://rework.withgoogle.com/en/guides/unbiasing-raise-awareness#introduction

Goretzki, L., Strauss, E., & Weber, J. (2013). An institutional perspective on the changes in management accountants' professional role. *Management Accounting Research*, *24*(1), 41–63. https://doi.org/10.1016/j.mar.2012.11.002

Goto, S. (2007). The Bounds of Classical Risk Management and the Importance of a Behavioral Approach. *Risk Management and Insurance Review*, *10*(2), 267–282. https://doi.org/10.1111/j.1540-6296.2007.00118.x

Graf, L., König, A., Enders, A., & Hungenberg, H. (2012). Debiasing competitive irrationality: How managers can be prevented from trading off absolute for relative profit. *European Management Journal*, *30*(4), 386–403. https://doi.org/10.1016/j.emj.2011.12.001

Hardies, K., Ohlrogge, F., Mentens, J., & Vandennieuwenhuysen, J. (2024). A Guide for Accounting Researchers to Conduct and Report Systematic Literature Reviews. *Behavioral Research in Accounting*, *36*(1), 21–43. https://doi.org/10.2308/BRIA-2022-042

Healy, M., & McCutcheon, M. (2008). Engagement With Active Learning: Reflections on the Experiences of Irish Accounting Students. *Accounting, Finance & Governance Review*, *15*(1). https://doi.org/10.52399/001c.26969

Hernik, J., & Jaworska, E. (2018). THE EFFECT OF ENJOYMENT ON LEARNING. In *INTED Proceedings.* IATED. https://doi.org/10.21125/inted.2018.1087

Hirsch, B. (2007). *Controlling und Entscheidungen: Zur verhaltenswissenschaftlichen Fundierung des Controllings*. *Die Einheit der Gesellschaftswissenschaften: Bd. 139*. Mohr Siebeck.

Hutchinson, J. W., Alba, J. W., & Eisenstein, E. M. (2010). Heuristics and Biases in Data-Based Decision Making: Effects of Experience, Training, and Graphical Data Displays. *Journal of Marketing Research*, *47*(4), 627–642. https://doi.org/10.1509/jmkr.47.4.627

Ichheiser, G. (1943). Misinterpretations of personality in everyday life and the psychologist's frame of reference. *Journal of Personality*, *12*(2), 145–160. https://doi.org/10.1111/j.1467-6494.1943.tb01953.x

Intelligence Advanced Research Projects Activity. (2012). *Unbiasing your biases I*. 522 Productions.

Intelligence Advanced Research Projects Activity. (2013). *Unbiasing your biases II*. 522 Productions.

Järvenpää, M. (2007). Making Business Partners: A Case Study on how Management Accounting Culture was Changed. *European Accounting Review*, *16*(1), 99–142. https://doi.org/10.1080/09638180701265903

Jensen, M., Lee, Y.-H., Piercy, C., Dunbar, N., Elizondo, J., Bessarabova, E., Twyman, N., Burgoon, J., Valacich, J., Adame, B., Miller, C., & Wilson, S. (2016). Exploring Failure and Engagement in a Complex Digital Training Game: A Multi-method Examination. *AIS Transactions on Human-Computer Interaction*, *8*(1), 1–20. https://aisel.aisnet.org/thci/vol8/iss1/2

Jugnandan, S., & Willows, G. D. (2023). Towards an integrated debiasing framework for consumer financial decisions: A reflection on debiasing research. *International Journal of Consumer Studies*, *47*(4), 1544–1560. https://doi.org/10.1111/ijcs.12935

Kahneman, D [Daniel]. (2011). *Thinking, fast and slow* (1st pbk. ed.). Farrar, Straus and Giroux.

Kahneman, D [Daniel], & Tversky, A [Amos] (1972). Subjective Probability: A Judgment of Representativeness. *Cognitive Psychology*, *3*(3), 430–454. https://doi.org/10.1016/0010-0285(72)90016-3

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *The Journal of Applied Psychology*, *90*(4), 677–691. https://doi.org/10.1037/0021-9010.90.4.677

Kirchgässner, G. (2013). *Homo Oeconomicus: Das ökonomische Modell individuellen Verhaltens und seine Anwendung in den Wirtschafts- und Sozialwissenschaften* (4., erg. und aktualisierte Aufl.). *Die Einheit der Gesellschaftswissenschaften: Bd. 74*. Mohr Siebeck.

Knowles, M. S. (1990). *The Adult Learner: A Neglected Species* (4th ed.). Gulf Publishing Company.

Kolb, D. A. (2015). *Experiential learning: Experience as the source of learning and development* (Second edition). Pearson Education LTD.

Korteling, J. E. H., Gerritsma, J. Y. J., & Toet, A. (2021). Retention and Transfer of Cognitive Bias Mitigation Interventions: A Systematic Literature Study. *Frontiers in Psychology*, *12*, Article 629354, 1–13. https://doi.org/10.3389/fpsyg.2021.629354

Kovic, M. (2019). Debiasing in a Minute or Less, or Too Good to Be True? The Effect of Micro-Interventions on Decision-Making Quality. *Psych*, *1*(1), 220–239. https://doi.org/10.3390/psych1010016

Kreilkamp, N., Schmidt, M., & Wöhrmann, A. (2019). Effizienzsteigerung durch Debiasing: Empfehlungen für die Praxis. *Controlling*, *31*(2), 57–65.

Kreilkamp, N., Schmidt, M., & Wöhrmann, A. (2021). Debiasing as a powerful management accounting tool? Evidence from German firms. *Journal of Accounting & Organizational Change*, *17*(4), 555–582. https://doi.org/10.1108/JAOC-12-2019-0122

Kucera, T. (2020). *Cognitive bias mitigation: How to make decision-making rational?* (IES Working Paper 01/2020). Prague: Charles University in Prague, Institute of Economic Studies (IES). https://www.econstor.eu/handle/10419/228083

Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 316–337). Blackwell Publishing Ltd.

Lee, Y.-H., Dunbar, N. E., Miller, C. H., Lane, B. L., Jensen, M. L., Bessarabova, E., Burgoon, J. K., Adame, B. J., Valacich, J. J., Adame, E. A., Bostwick, E., Piercy, C. W., Elizondo, J., & Wilson, S. N. (2016). Training Anchoring and Representativeness Bias Mitigation Through a Digital Game. *Simulation & Gaming*, *47*(6), 751–779. https://doi.org/10.1177/1046878116662955

Legaki, N.-Z., Karpouzis, K., Assimakopoulos, V., & Hamari, J. (2021). Gamification to avoid cognitive biases: An experiment of gamifying a forecasting course. *Technological Forecasting and Social Change*, *167*, 120725. https://doi.org/10.1016/j.techfore.2021.120725

Li, G., Cai, X., Yang, Q., Cui, Q., Huang, L., Jing, X., & Wang, Y. (2023). A review of attentional bias modification trainings for depression. *CNS Neuroscience & Therapeutics*, *29*(3), 789–803. https://doi.org/10.1111/cns.14022

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. https://doi.org/10.1016/j.mar.2018.09.002

Luong, R., & Butler, K. (2022). Informal psychology education and the fundamental attribution error: Testing the effectiveness of accessible online passages. *Current Psychology*, 1–13. https://doi.org/10.1007/s12144-021-02432-w

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

Malone, T. W. (1981). Toward a Theory of Intrinsically Motivating Instruction *Cognitive Science*, *5*(4), 333–369. https://doi.org/10.1207/s15516709cog0504_2

Martey, R. M [Rosa M.], Shaw, A., Stromer-Galley, J., & Kenski, K. (2017). Testing the Power of Game Lessons: The Effects of Art Style and Narrative Complexity on Reducing Cognitive Bias. *International Journal of Communication*, *11*, 1636–1660.

McKernan, B., Martey, R. M [Rosa Mikeal], Stromer-Galley, J., Kenski, K., Clegg, B. A., Folkestad, J. E., Rhodes, M. G., Shaw, A., Saulnier, E. T [Emilie T.], & Strzalkowski, T. (2015). We don't need no stinkin' badges: The impact of reward features and feeling rewarded in educational games. *Computers in Human Behavior*, *45*, 299–306. https://doi.org/10.1016/j.chb.2014.12.028

Mersch, E., Fox, O., Leonard, J [Jim], Flach, J., Bennett, K., & Holt, J. (2013). Mitigation of the Confirmation Bias Using a Game-Based Trainer. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 1475–1479. https://doi.org/10.1177/1541931213571329

Montibeller, G., & Winterfeldt, D. von (2015). Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis: An Official Publication of the Society for Risk Analysis*, *35*(7), 1230–1251. https://doi.org/10.1111/risa.12360

Moore, D. A. (2023). Overprecision is a property of thinking systems. *Psychological Review*, *130*(5), 1339–1350. https://doi.org/10.1037/rev0000370

Morewedge, C. K., & Kahneman, D [Daniel] (2010). Associative processes in intuitive judgment. *Trends in Cognitive Sciences*, *14*(10), 435–440. https://doi.org/10.1016/j.tics.2010.07.004

Morewedge, C. K., Yoon, H., Scopelliti, I [Irene], Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015). Debiasing Decisions: Improved Decision Making With a Single Training Intervention. *Policy Insights from the Behavioral and Brain Sciences*, *2*(1), 129– 140. https://doi.org/10.1177/2372732215600886

Mullinix, G., Gray, O., Colado, J., Veinott, E., Leonard, J [James], Papautsky, E. L., Argenta, C., Clover, M., Sickles, S., Castronova, E., Todd, P. M., Ross, T., Lorince, J., Hoteling, J., Mayell, S., Hale, C [Chris], Whitaker, E., Hoffman, R. R., Fox, O., & Flach, J. (2013). Heuristica: Designing a serious game for improving decision making. *2013 IEEE International Games Innovation Conference (IGIC)*, 250–255. https://doi.org/10.1109/IGIC.2013.6659159

Murata, A. (2016). Discussion on Feasibility of Removal of Cognitive Biases. *Journal of Psychology Research*, *6*(7). https://doi.org/10.17265/2159-5542/2016.07.001

Neilens, H. L., Handley, S. J., & Newstead, S. E. (2009). Effects of training and instruction on analytic and belief-based reasoning processes. *Thinking & Reasoning*, *15*(1), 37–68. https://doi.org/10.1080/13546780802535865

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

O'Neil, H., & Chen, H.-H. (2005). A formative evaluation of the training effectiveness of a computer game. https://www.semanticscholar.org/paper/A-formative-evaluation-of-the-training-of-a-game-O%27Neil-Chen/973152cb3cbb9075681302ca385dc4ff69044f0a

Peters, R. A. (2011). Enhancing academic achievement by identifying and minimizing the impediments to active learning. *Public Administration Quarterly*, *35*(4), 466–493. http://www.jstor.org/stable/23209325

Poos, J. M., van den Bosch, K., & Janssen, C. P. (2017). Battling bias: Effects of training and training context. *Computers & Education*, *111*, 101–113. https://doi.org/10.1016/j.compedu.2017.04.004

Pronin, E., Lin, D. Y., & Ross, L. (2002). The Bias Blind Spot: Perceptions of Bias in Self Versus Others. *Personality & Social Psychology Bulletin*, *28*(3), 369–381. https://doi.org/10.1177/0146167202286008

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology*, *15*, 23–32. https://doi.org/10.1002/1099-0720

Rhodes, R. E., Kopecky, J., Bos, N., McKneely, J., Gertner, A., Zaromb, F., Perrone, A., & Spitaletta, J. (2017). Teaching Decision Making With Serious Games. *Games and Culture*, *12*(3), 233–251. https://doi.org/10.1177/1555412016686642

Riley, J., & Ward, K. (2017). Active Learning, Cooperative Active Learning, and Passive Learning Methods in an Accounting Information Systems Course. *Issues in Accounting Education*, *32*(2), 1–16. https://doi.org/10.2308/iace-51366

Robbins, J. M., & Krueger, J. I. (2005). Social Projection to Ingroups and Outgroups: A Review and Meta-Analysis. *Personality and Social Psychology Review*, *9*(1), 32–47. https://doi.org/10.1207/s15327957pspr0901_3

Rooney, P. (2012). A Theoretical Framework for Serious Game Design. *International Journal of Game-Based Learning*, *2*(4), 41–60. https://doi.org/10.4018/ijgbl.2012100103

Rumeser, D., & Emsley, M. (2019). Can Serious Games Improve Project Management Decision Making Under Complexity? *Project Management Journal*, *50*(1), 23–39. https://doi.org/10.1177/8756972818808982

Sandberg, S. (2015). *Managing Unconscious Bias*. https://about.fb.com/news/2015/07/managing-unconscious-bias/

Schäffer, U. (2013). Management accounting research in Germany: From splendid isolation to being part of the international community. *Journal of Management Control*, *23*(4), 291–309. https://doi.org/10.1007/s00187-013-0170-6

Scopelliti, I [Irene]. (2023). Training Can Improve Decision Making. In J. Musolino, J. Sommer, & P. Hemmer (Eds.), *The cognitive science of belief: A multidisciplinary approach* (pp. 557–573). Cambridge University Press. https://doi.org/10.1017/9781009001021.039

Sellier, A.-L., Scopelliti, I [Irene], & Morewedge, C. K. (2019a). Debiasing Training Improves Decision Making in the Field. *Psychological Science*, *30*(9), 1371–1379. https://doi.org/10.1177/0956797619861429

Sellier, A.-L., Scopelliti, I [I.], & Morewedge, C. K. (2019b). New Evidence Reveals Training Can Reduce Cognitive Bias And Improve Decision-making. *Forbes*. https://www.forbes.com/sites/hecparis/2019/10/08/new-evidence-reveals-training-can-reduce-cognitive-bias--and-improve-decision-making/

Shaw, A., Kenski, K., Stromer-Galley, J., Mikeal Martey, R., Clegg, B. A., Lewis, J. E., Folkestad, J. E., & Strzalkowski, T. (2018). Serious Efforts at Bias Reduction. *Journal of Media Psychology*, *30*(1), 16–28. https://doi.org/10.1027/1864-1105/a000174

Shepperd, M., Mair, C., & Jørgensen, M. (2018). An experimental evaluation of a de-biasing intervention for professional software developers. In H. M. Haddad, R. L. Wainwright, & R. Chbeir (Eds.), *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (pp. 1510–1517). ACM. https://doi.org/10.1145/3167132.3167293

Simon, H. A. (1978). Rationality as Process and as Product of Thought. *The American Economic Review*, *68*(2), 1–16.

Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology*, *64*(2), 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The Comparative Effectiveness of Web-Based and Classroom Instruction: A Meta-Analysis. *Personnel Psychology*, *59*(3), 623–664. https://doi.org/10.1111/j.1744-6570.2006.00049.x

Snowden, M., & Halsall, J. P. (2016). Self-determined approach to learning: A social science perspective. *Cogent Education*, *3*(1), 1247608. https://doi.org/10.1080/2331186X.2016.1247608

Soll, J. B., Milkman, K. L., & Payne, J. W. (2015). A User's Guide to Debiasing. In G. Keren & G. Wu (Eds.), *The Wiley-Blackwell Handbook of Judgment and Decision making.* Wiley-Blackwell.

Stanovich, K. E. (2011). *Rationality and the reflective mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195341140.001.0001

Stapleton, A. J. (2004). Serious Games: Serious Opportunities. *Paper Presented at the Australian Game Developers' Conference*.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Symborski, C., Barton, M., Quinn, M. M., Korris, J. H., Kassam, K. S., & Morewedge, C. K. (2017). The Design and Development of Serious Games Using Iterative Evaluation. *Games and Culture*, *12*(3), 252–268. https://doi.org/10.1177/1555412016673262

Symborski, C., Barton, M., Quinn, M., Morewedge, C. K., & Kassam, K. S. (2014). Missing: A Serious Game for the Mitigation of Cognitive Biases. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Article 14295, 1–13.

Thaler, R. (1985). Mental Accounting and Consumer Choice. *Marketing Science*, *4*(3), 199–214. http://www.jstor.org/stable/183904

Tommasi, F., Ceschi, A., Weller, J., Costantini, A., Passaia, G., Gostimir, M., & Sartori, R. (2021). An empirical evaluation of tech interventions to improve financial decision-making. *European Journal of Training and Development*, *45*(6/7), 633–649. https://doi.org/10.1108/EJTD-11-2020-0169

Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*(14), 207–222.

Tversky, A [A.], & Kahneman, D [D.] (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Tversky, A [A.], & Kahneman, D [Daniel] (1981). The Framing of Decisions and the Psychology of Choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

van Brussel, S., Timmermans, M., Verkoeijen, P., & Paas, F. (2021). Teaching on video as an instructional strategy to reduce confirmation bias—a pre-registered study. *Instructional Science*, *49*(4), 475–496. https://doi.org/10.1007/s11251-021-09547-4

Veinott, E. S., Leonard, J [James], Papautsky, E. L., Perelman, B., Stankovic, A., Lorince, J., Hotaling, J., Ross, T., Todd, P., Castronova, E., Busemeyer, J., Hale, C [Christoper], Catrambone, R., Whitaker, E., Fox, O., Flach, J., & Hoffman, R. R. (2013). The Effect of Camera Perspective and Session Duration on Training Decision Making in a Serious Video Game. *2013 IEEE International Games Innovation Conference (IGIC)*, 256–262. https://doi.org/10.1109/IGIC.2013.6659170

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). Contributions of Specific Cell Information to Judgments of Interevent Contingency. *Journal of Experimental Psychology*, *16*(3), 509–521. https://doi.org/10.1037//0278-7393.16.3.509

Weinstein, N. D., & Klein, W. M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychology : Official Journal of the Division of Health Psychology, American Psychological Association*, *14*(2), 132–140. https://doi.org/10.1037//0278-6133.14.2.132

Welsh, M. B., Begg, S. H., & Bratvold, R. B. (2007). Efficacy of bias awareness in debiasing oil and gas judgments. *Proceedings of the Annual Meeting of the Cognitive Science Society*(29), 1647–1652. https://escholarship.org/content/qt6qx02997/qt6qx02997.pdf

Wilson, K. A., Bedwell, W. L., Lazzara, E. H., Salas, E., Burke, C. S., Estock, J. L., Orvis, K. L., & Conkey, C. (2009). Relationships Between Game Attributes and Learning Outcomes. *Simulation & Gaming*, *40*(2), 217–266. https://doi.org/10.1177/1046878108321866

Yoon, H., Scopelliti, I [Irene], & Morewedge, C. K. (2021). Decision making can be improved through observational learning. *Organizational Behavior and Human Decision Processes*, *162*, 155–188. https://doi.org/10.1016/j.obhdp.2020.10.011

TABLE 1
Definitions of investigated cognitive biases

Bias classification	Cognitive bias	Definition
Motivational	*Confirmation Bias*	Tendency to select and process preferably information that confirms the decision-maker's existing opinion and hypotheses (Nickerson, 1998).
	Fundamental Attribution Error	Tendency to attribute a person's behavior rather to his personality than to the situation (Gilbert, 1998; Ichheiser, 1943).
	Blind Spot Bias	Tendency to perceive one's own decisions as less biased than the decisions of others (Pronin et al., 2002).
	Projection Bias	Assumption that emotions, values, and preferences of others are similar to one's own (Robbins & Krueger, 2005).
Statistical	*Representativeness Bias*	The probability of an event is estimated based on its similarity to a representative event instead of its statistical probability (Kahneman & Tversky, 1972).
	*Insensitivity to Sample Size**	Tendency to ignore that, when estimating probabilities, characteristics of small samples differ from those of bigger samples (for the same population; especially relevant for variance) (Tversky & Kahneman, 1974).
	*Base Rate Neglect**	Tendency to ignore the base rate of an event when estimating probabilities of related events (Tversky & Kahneman, 1974).
	*Stereotype Bias**	Tendency to make predictions solely based on the applicability of a stereotypic description that is assumed to be accurate (Lee et al., 2016).
	*Conjunction Fallacy**	Tendency to assess the probability of the conjunction of two events higher than the probability of each elementary event (Tversky & Kahneman, 1974).
	*Gambler's Fallacy**	Tendency to expect that a sequence of events resulting from a random process will accurately reflect that process's characteristics (i.e., randomness) (Kahneman, 2011; Lee et al., 2016).
	*Insensitivity to Predictability**	Tendency to rely on favorable descriptions rather than relevant information for predictions (Tversky & Kahneman, 1974).
	Regression to the Mean	Tendency to neglect that a value following an extreme value (statistical outlier) will most presumably be closer to the mean value (Barnett et al., 2005).
	Covariation Detection	Tendency to ignore elements and information from the contingency table when judging whether statistical dimension A influences dimension B. Instead of calculating the effect statistically, one only considers which combination has the highest amount of observations (Wasserman et al., 1990).
Informational	*Anchoring Effect*	Tendency to anchor estimates too much or without sufficient adaption to initial or existing reference values. As a consequence, biased estimates result (Tversky & Kahneman, 1974).
	Ambiguity Aversion	Tendency to avoid ambiguous information coupled with low probability of outcomes in decision-making (Fox & Tversky, 1995; Tommasi et al., 2021).
	Information Bias	Tendency to seek as much information as possible, although the extra information is irrelevant and does not lead to a better decision (Baron et al., 1988; Tommasi et al., 2021).
	Mental Accounting Error	Tendency to assign different values on the same amount of money based on subjective criteria (Thaler, 1985; Tommasi et al., 2021).

Overconfidence Bias	Tendency to overestimate one's own abilities and knowledge (Arkes, 1991).
*Overprecision Bias***	Tendency of being excessively confident of the accuracy of one's judgement (Moore, 2023).
Framing Effect	Tendency to come to different decisions depending on the way identical information is presented/framed (Tversky & Kahneman, 1981).
Outcome Bias	Tendency to rate a decision based on its outcome and consequences rather than considering the underlying decision process (Baron & Hershey, 1988).
Sunk Cost Fallacy	Tendency and error to consider past costs (i.e., sunk costs) and time invested for future decisions (Arkes & Blumer, 1985).

Notes:
* These biases are often considered as subgroups of the representation bias.
** Overprecision is one (out of three) forms of the overconfidence bias.

TABLE 2
List of studies included in the review

#	Study	Year	Training approach
[1]	**Aczel**, B./Bago, B./Szollosi, A./Foldes, A./Lukacs, B.	2015	Analogical Training; Awareness Training
[2]	**Adame**, B. J.	2016	Awareness Training
[3]	**AlKhars**, M./Evangelopoulos, N./Pavur, R./Kulkarni, S.	2019	Awareness Training
[4]	**Barton**, M./Symborski, C./Quinn, M./Morewedge, C. K./Kassam, K. S./Korris, H. K.	2016	Serious Game; Training Video
[5]	**Bessarabova**, E./Piercy, C. W./King, S./Vincent, C./Dunbar, N. E./Burgoon, J. K./Miller, C. H./Jensen, M./Elkins, A./Wilson, D. W./Wilson, S. N./Lee, Y.-H.	2016	Serious Game
[6]	**Boissin**, E./Caparos, S./Voudouri, A./De Neys, W.	2022	Awareness Training
[7]	**Cassotti**, M./Moutier, S.	2010	Awareness Training
[8]	**Clegg**, B. A./Martey, R. M./Stromer-Galley, J./Kenski, K./Saulnier, E. T./Folkestad, J. E./McLaren, E./Shaw, A./Lewis, J. E./Patterson, J. D./Strzalkowski, T.	2014	Serious Game; Training Video
[9]	**Clegg**, B. A./McKernan, B./Martey, R. M./Taylor, S. M./Stromer-Galley, J./Kenski, K./Saulnier, E. T./Rhodes, M. G./Folkestad, J. E./McLaren, E./Shaw, A./Strzalkowski, T.	2015	Serious Game; Training Video
[10]	**Dunbar**, N. E./Jensen, M. L./Miller, C. H./Bessarabova, E./Lee, Y.-H./Wilson, S. N./Elizondo, J./Adame, B. J./Valacich, J. S./Straub, S. K./Burgoon, J. K./Lane, B. L./Piercy, C. W./Wilson, D./King, S./Vincent, C./Schuetzler, R. M.	2017	Serious Game; Training Video
[11]	**Dunbar**, N. E./Jensen, M. L./Miller, C. H./Bessarabova, E./Straub, S. K./Wilson, S. N./Elizondo, J./Burgoon, J. K./Valacich, J. S./Adame, B./Lee, Y.-H./Lane, B. L./Piercy, C./Wilson, D./King, S./Vincent, C./Scheutzler, R.	2014	Serious Game; Training Video
[12]	**Dunbar**, N. E./Miller, C. H./Adame, B. J./Elizondo, J./Wilson, S. N./Lane, B. L./Kauffman, A. A./Bessarabova, E./Jensen, M. L./Straub, S. K./Lee, Y.-H./Burgoon, J. K./Valacich, J. S./Jenkins, J./Zhang, J.	2014	Serious Game; Training Video
[13]	**Lee**, Y.-H./Dunbar, N. E./Miller, C. H./Lane, B. L./Jensen, M. L./Bessarabova, E./Burgoon, J. K./Adame, B. J./Valacich, J. J./Adame, E. A./Bostwick, E./Piercy, C. W./Elizondo, J./Wilson, S. N.	2016	Serious Game; Awareness Training
[14]	**Luong**, R./Butler, K.	2022	Awareness Training
[15]	**Martey**, R. M./Shaw, A./Stromer-Galley, J./Kenski, K./Clegg, B. A./Folkestad, J. E./Saulnier, E. T./Strzalkowski, T.	2017	Serious Game; Training Video
[16]	**McKernan**, B./Martey, R. M./Stromer-Galley, J./Kenski, K./Clegg, B. A./Folkestad, J. E./Rhodes, M. G./Shaw, A./Saulnier, E. T./Strzalkowski, T.	2015	Serious Game
[17]	**Morewedge**, C. K./Yoon, H./Scopelliti, I./Symborski, C. W./Korris, J. H./Kassam, K. S.	2015	Serious Game; Training Video
[18]	**Moutier**, S./Houdé, O.	2003	Awareness Training

TABLE 2 (continued)

[19]	**Poos**, J. M./van den Bosch, K./Janssen, C. P.	2017	Serious Game; Reading Serious Game Script
[20]	**Rhodes**, R.E./Kopecky, J./Bos, N./McKneely, J./Gertner, A./Zaromb, F./Perrone, A./Spitaletta, J.	2017	Serious Game; Training Video
[21]	**Scopelliti**, I./Morewedge, C. K./McCormick, E./Min, H. L./Lebrecht, S./Kassam, K. S.	2015	Awareness Training
[22]	**Sellier**, A.-L./Scopelliti, I./Morewedge, C. K.	2019	Serious Game
[23]	**Shaw**, A./Kenski, K./Stromer-Galley, J./Martey, R. M./Clegg, B. A./Lewis, J. E./Folkestad, J. E./Strzalkowski, T.	2018	Serious Game; Training Video
[24]	**Shepperd**, M./Mair, C./Jørgensen, M.	2018	Awareness Training
[25]	**Symborski**, C. W./Barton, M./Quinn, M./Morewedge, C. K./Kassam, K. S./Korris, J. H.	2014	Serious Game; Training Video
[26]	**Tommasi**, F./Ceschi, A./Weller, J./Costantini, A./Passaia, G./Gostimir, M./Sartori, R.	2021	Serious Game
[27]	**van Brussel**, S./Timmermans, M./Verkoeijen, P./Paas, F.	2021	Awareness Training
[28]	**Veinott**, E./Leonard, J./Papautsky, E. L./Perelman, B./Stankovic, A./Lorince, J./Hotaling, J./Ross, T./Mayell, S./Todd, P. M./Busemeyer, J./Castronova, E./Hale, C./Catrambone, R./Whitaker, E./Fox, O./Flach, J./Hoffman, R. R./	2013	Serious Game; Training Video
[29]	**Welsh**, M. B./Begg, S. H./Bratvold, R. B.	2007	Awareness Training
[30]	**Yoon**, H./Scopelliti, I./Morewedge, C. K.	2021	Observational Training; Serious Game; Training Video

TABLE 3
Overview of the effectiveness of the training interventions included in the review

Biases			Training approaches						Effectiveness - Comparison of training approaches						Long-term effectiveness				Transfer	Studies
Bias classification	*Bias*	*Number of Studies*	*Serious Game*	*Training Video*	*Reading Serious Game Script*	*Observational Training*	*Awareness Training*	*Analogical Training*	*Game vs. Video*	*Game vs. Reading Game Script*	*Game vs. Observation*	*Observation vs. Video*	*Game vs. Awareness*	*Analogical vs. Awareness*	*Serious Games*	*Training Video*	*Awareness Training*	*Analogical Training*	*Knowledge Transfer*	
Motivational	Confirmation bias	13	-	-	-		-		G > V	G = R					yes	yes			yes	[8], [10], [11], [12], [15], [17], [19], [20], [22], [23], [25], [27], [28]
	Fundamental attribution error	13	-	-	-		o/-		G ≥ V	G = R					yes	yes			no	[8], [10], [11], [12], [14], [15], [17], [19], [20], [21], [23], [25], [28]
	Blind spot bias	8	-	-					G ≥ V						yes	yes				[5], [8], [15], [17], [20], [23], [25], [28]
	Projection bias	6	-	-		-			G > V		G > O	O > V			yes	yes				[4], [9], [16], [17], [20], [30]
Statistical	Representativeness bias*	7	-	-		-			G ≥ V		G = O	O = V			yes	yes				[4], [9], [16], [17], [20], [26], [30]
	Insensitivity to sample size	2	-				o								no					[3], [13]
	Base rate neglect	3	-				-						G = A		no		yes			[3], [6], [13]
	Stereotype bias	1	-				-						G = A		no		no			[13]
	Conjunction fallacy	4	-				-										yes			[6], [7], [18], [26]
	Gambler's fallacy	2	o				o/-													[3], [13]
	Insensitivity to predictability	1					-													[3]
	Regression to the mean	1					o													[3]
	Composite statistical bias**	1					o	-						An ≥ A			no	yes	yes	[1]

TABLE 3 (continued)

Biases		Training approaches							Effectiveness - Comparison of training approaches						Long-term effectiveness				Transfer	Studies
Bias classification	*Bias*	*Number of Studies*	*Serious Game*	*Training Video*	*Reading Serious Game Script*	*Observational Training*	*Awareness Training*	*Analogical Training*	*Game vs. Video*	*Game vs. Reading Game Script*	*Game vs. Observation*	*Observation vs. Video*	*Game vs. Awareness*	*Analogical vs. Awareness*	*Serious Games*	*Training Video*	*Awareness Training*	*Analogical Training*	*Knowledge Transfer*	
Informational	Anchoring	12	-	-		-	o/-	o	G ≥ V		G = O	O = V			yes	yes				[4], [9], [20], [17], [16], [13], [30], [1], [26], [2], [24], [29]
	Ambiguity aversion	1	o																	[26]
	Information bias	1	o																	[26]
	Mental accounting error	1	-																	[26]
	Overconfidence bias	2					o/-	o												[1], [29]
	Overprecision bias***	1					o													[3]
	Framing effect	1					o	o												[1]
	Outcome bias	1					o	o												[1]
	Sunk cost fallacy	1					o	o												[1]
Number of studies		**30**	**19**	**13**	**1**	**1**	**12**	**1**	**13**	**1**	**1**	**1**	**1**	**1**	**12**	**10**	**2**	**1**	**3**	[1] – [30]

Notes:
* The representativeness bias is often measured by combining items that measure the base rate neglect, conjunction fallacy, insensitivity to sample size and gambler's fallacy.
** The composite statistical bias consists of the following biases: base rate neglect, insensitivity to sample size, regression to the mean, and covariation detection.
*** Overprecision is one (of three) form(s) of overconfidence.

Symbol	Meaning
=	equally/comparably effective as
≥	generally more effective
>	more effective than
-	significant reducing effect
o	no reducing effect
G	Serious Game
V	Training Video
R	Reading Serious Game Script
O	Observational Training
A	Awareness Training
An	Analogical Training

TABLE 4
Effectiveness of the training interventions on bias recognition and discrimination

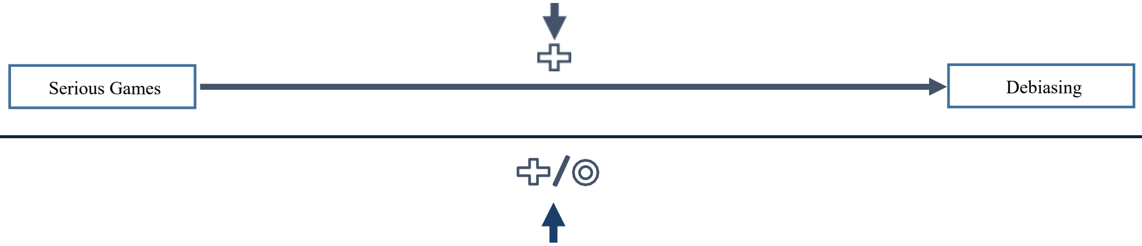
Biases		Training interventions				Effectiveness - comparison of training interventions			Long-term effectiveness		Studies
Bias classification	*Bias*	*Number of Studies*	*Serious Game*	*Training Video*	*Observational Training*	*Game vs. Video***	*Game vs. Observation*	*Observation vs. Video*	*Serious Games*	*Training Video*	
Motivational	Confirmation bias	6	+	+		G ≤ V			mainly yes	yes	[8], [12], [17], [20], [25], [28]
	Fundamental attribution error	6	+	+		G ≤ V			mainly yes	yes	[8], [12], [17], [20], [25], [28]
	Blind spot bias	6	+	+		G ≤ V			yes	yes	[5], [8], [17], [20], [25], [28]
	Projection bias	5	+	+	+	G < V	G = O	O < V	yes	yes	[4], [9], [16], [20], [30]
Statistical	Representa-tiveness bias*	6	+	+	+	G < V	G = O	O < V	yes	yes	[4], [9], [16], [17], [20], [30]
Informational	Anchoring	6	+	+	+	G < V	G = O	O < V	yes	yes	[4], [9], [16], [17], [20], [30]
Number of studies		**11**	**11**	**8**	**1**	**5**	**1**	**1**	**8**	**5**	[4], [5], [8], [9], [12], [16], [17], [20], [25], [28], [30]

Notes:
* The representativeness bias is often measured by combining items that measure the base rate neglect, conjunction fallacy, insensitivity to sample size and gambler's fallacy.
** The indicated results do not include the findings from Symborski et al. (2014) and Clegg et al. (2014 & 2015). Symborski et al. (2014) find that the game is superior to the video and Clegg et al. (2014 & 2015) find that the repeated game or both games taken together (repeated and single game play) outperform the video training.

Symbol	Meaning
=	equally/comparably effective as
≤	generally less effective than
<	less effective than
+	significant positive effect
G	Serious Game
V	Training Video
O	Observational Training

TABLE 5
Serious Game — Moderators

Training Approach	Effectiveness	Studies
Serious Games	Serious games reduce biases by explaining biases on an abstract level that requires personal engagement, which leads to enjoyment and promotes bias awareness, and by providing personalized feedback and debiasing strategies with immediate practice testing, which ultimately fosters bias knowledge and reduction.	Barton et al. (2015) / Bessarabova et al. (2016) / Clegg et al. (2014) / Clegg et al. (2015) / Dunbar et al. (2017) / Dunbar et al. (2014a) / Dunbar et al. (2014b) / Lee et al. (2016) / Martey et al. (2017) / McKernan et al. (2015) / Morewedge et al. (2015) / Poos et al. (2017) / Rhodes et al. (2017) / Sellier et al. (2019) / Shaw et al. (2018) / Symborski et al. (2014) / Veinott et al. (2013)



Moderating effects

Moderator	Operationalization	Prediction	Bias	Studies[a] & Effectiveness: Yes	Studies[a] & Effectiveness: No
Duration	Longer game duration	▪ A longer game duration allows players to become more familiar with the game mechanics, which reduces the cognitive load associated with learning a new game and allows their focus to shift to the actual training content *(Dunbar et al. 2014a).* ▪ Extended exposure to the game enhances retention of the learning material.	▪ Confirmation bias (1/4) ▪ Fundamental attribution error (1/4) ▪ Blind spot bias (0/2)	*Dunbar et al. (2014a):* Confirmation bias, Fundamental attribution error	*Dunbar et al. (2017) & Dunbar et al. (2014b):* Confirmation bias, Fundamental attribution error *Veinott et al. (2013):* Confirmation bias, Fundamental attribution error, Blind spot bias *Bessarabova et al. (2016):* Blind spot bias
Repetition	Playing the game twice	▪ Through repeated game play, individuals accumulate more knowledge about biases, which consequently leads to better bias mitigation.	▪ Confirmation bias (3/5) ▪ Fundamental attribution error (1/5) ▪ Blind spot bias (0/3) ▪ Projection bias (1*/1) ▪ Anchoring (1/1) ▪ Representativeness bias (1*/1)	*Clegg et al. (2014):* Confirmation bias *Clegg et al. (2015):* Projection bias *(*only when measured several weeks after the training),* Anchoring, Representativeness bias *(*only when measured several weeks after the training)* *Dunbar et al. (2017):* Confirmation bias *Dunbar et al. (2014a):* Confirmation bias, Fundamental attribution error	*Bessarabova et al. (2016):* Blind spot bias *Clegg et al. (2014):* Fundamental attribution error, Blind spot bias *Dunbar et al. (2014b):* Confirmation bias, Fundamental attribution error *Dunbar et al. (2017):* Fundamental attribution error *Veinott et al. (2013):* Confirmation bias, Fundamental attribution error, Blind spot bias

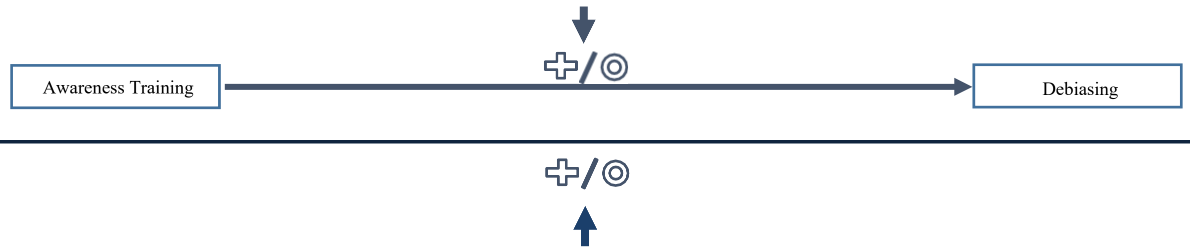
TABLE 5 (continued)

Moderator	Operationalization	Prediction	Bias	Studies[a] & Effectiveness: Yes	Studies[a] & Effectiveness: No
		Moderating effects			
Player type	Multiplayer *(i.e., playing the game in a group of two)*	▪ Making decisions in a group requires each team member to be prepared to defend their reasoning to the group, which helps to reduce biases. ▪ Collaborative multiplayer gaming provides the opportunity to analyze alternative sources of feedback and increases the understanding of abstract concepts. ▪ Cooperative learning improves learning outcomes.	▪ Confirmation bias (0/2) ▪ Fundamental attribution error (0/2) ▪ Blind spot bias (0/1)		*Bessarabova et al. (2016):* Blind spot bias *Dunbar et al. (2014a) & Dunbar et al. (2017):* Confirmation bias *(single-player mode is more effective),* Fundamental attribution error
Feedback timing	Immediate in game feedback	▪ Immediate feedback during the game leads to faster learning, which allows to correct mistakes, adjust decisions and thus improve performance, compared to delayed feedback after task completion.	▪ Confirmation bias (0/1) ▪ Fundamental attribution error (0/1) ▪ Blind spot bias (0/1)		*Dunbar et al. (2017):* Confirmation bias, Fundamental attribution error *Bessarabova et al. (2016):* Blind spot bias
Game style	Choosing own avatar design	▪ Customizing an avatar (i.e., a player character) improves players' involvement in the game, as they identify more strongly with their avatar, which in turn leads to greater learning effects.	▪ Confirmation bias (0/1) ▪ Fundamental attribution error (0/1) ▪ Blind spot bias (0/1)		*Shaw et al. (2018):* Confirmation bias; Fundamental attribution error; Blind spot bias
	Detailed and realistic art style	▪ Rich and realistic visuals enhance feelings of co-presence and positively affect avatar credibility, which leads to greater engagement and learning outcomes.	▪ Confirmation bias (0/1) ▪ Fundamental attribution error (0/1) ▪ Blind spot bias (0/1)		*Martey et al. (2017):* Confirmation bias, Fundamental attribution error; Blind spot bias (*minimalistic art style is slightly superior*)
	Rich narrative *(i.e., named characters with backstories, plot events, broader world context)*	▪ Rich narratives foster self-reflection, encourage exploratory learning, offer captivating learning experiences, and enhance involvement and knowledge transfer, which leads to increased learning results (Martey et al. 2017). ▪ A rich narrative provides additional depth of comprehension, enabling learners to connect instructional content with familiar, ultimately enhancing the transfer of learning (Martey et al. 2017).	▪ Confirmation bias (0/1) ▪ Fundamental attribution error (0/1) ▪ Blind spot bias (0/1)		*Martey et al. (2017):* Confirmation bias (*marginal significance for minimal narrative style*); Fundamental attribution error; Blind spot bias
	3rd person camera perspective	▪ Changing the camera perspective from 1st perspective (i.e., being the avatar) to the 3rd perspective (i.e., being behind the avatar) shift players away from a self-oriented frame of reference and enables reorientation towards the environment, which improves bias mitigation.	▪ Confirmation bias (0/1) ▪ Fundamental attribution error (0/1) ▪ Blind spot bias (0/1)		*Veinott et al. (2013):* Fundamental attribution error; Blind spot bias; Confirmation bias
	Explicit training	▪ Enhancing the gameplay (i.e., implicit training) with additional bias definitions and knowledge questions (i.e., explicit/hybrid training) reduces the cognitive load required to comprehend the learning materials and enables trainees to more effectively assimilate knowledge and apply the provided mitigation strategies.	▪ Confirmation bias (1/1) ▪ Fundamental attribution error (0/1) ▪ Blind spot bias (0/1)	*Dunbar et al. (2014b):* Confirmation bias	*Dunbar et al. (2014b):* Fundamental attribution error *Bessarabova et al. (2016):* Blind spot bias
Rewards	Granting a high number of rewards *(e.g., badges)*	▪ Rewards provide players with incentives to keep playing. ▪ A greater quantity of rewards leads to greater learning outcomes.	▪ Projection bias (0/1) ▪ Anchoring (0/1) ▪ Representativeness bias (0/1)		*McKernan et al. (2015):* Projection bias, Anchoring, Representativeness bias

+ = Positive(ly moderating) effect on debiasing | o = no moderating effect on debiasing.
(a/b) = The positively moderating effect is found in a out of b studies.
[a] Studies that examine moderating factors are presented in this table. All other relevant studies are listed above.

TABLE 6
Awareness Training — Moderators

Training Approach	Effectiveness	Studies
Awareness Training	A training intervention that presents information explaining the concept of cognitive biases through definitions, real-life examples of biased decision-making, strategies for bias mitigation, and practical applications.	Aczel et al. (2015); Boissin et al. (2022); Lee et al. (2016)

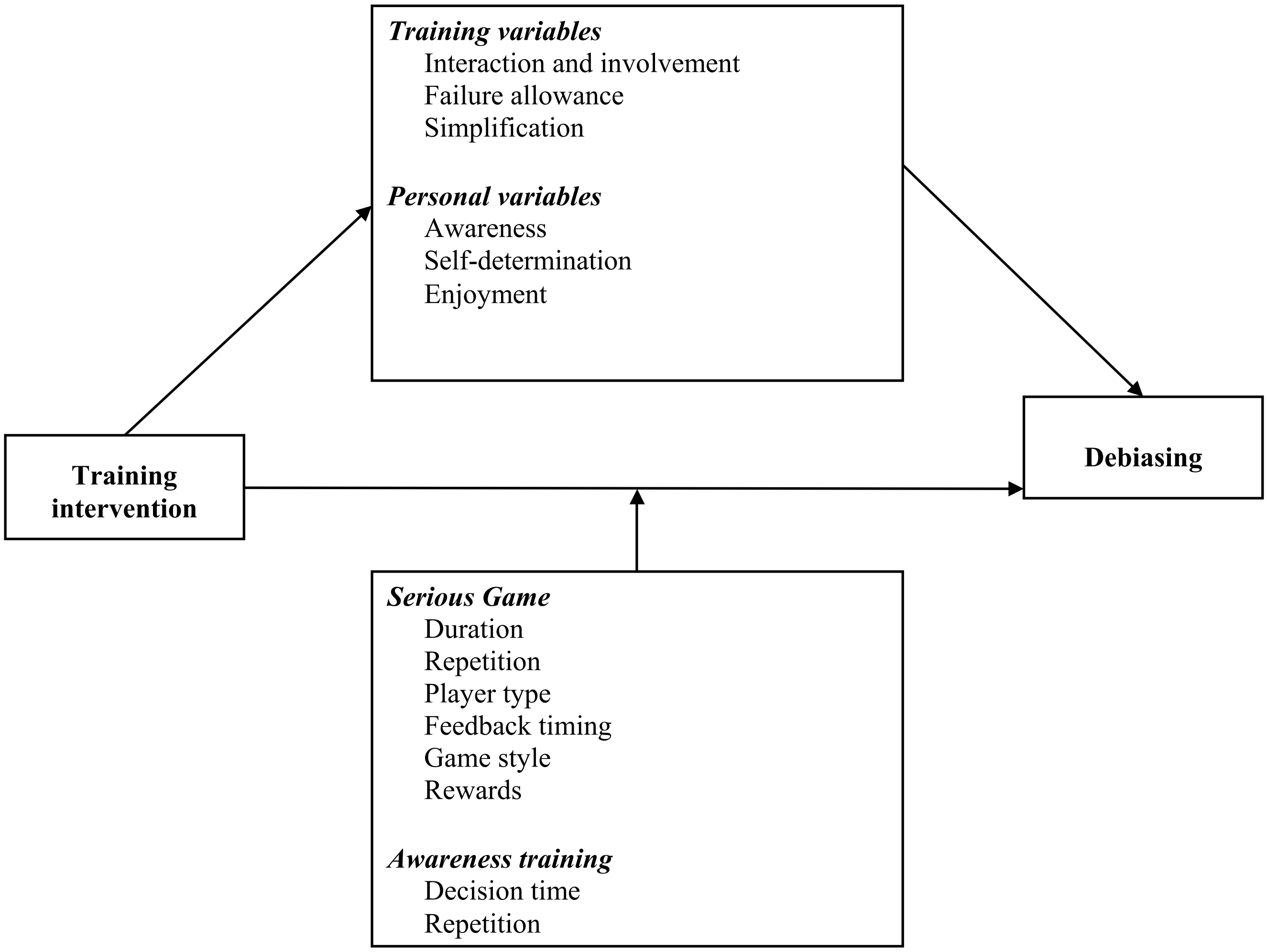


Moderating effects

Moderator	Operationalization	Prediction	Bias	Studies[a] & Effectiveness: Yes	Studies[a] & Effectiveness: No
Decision time	Intuitive (short)	▪ The provision of a short single-shot explanation of intuitive biases along with a correct solution strategy helps decision-makers to be rational. Following a thorough problem explanation, individuals solve structurally similar problems intuitively, even under time pressure.	▪ Base rate neglect (1/1) ▪ Conjunction fallacy (1/1)	*Boissin et al. (2022):* Base rate neglect, Conjunction fallacy	
	Deliberate (long)	▪ In the absence of time pressure and/or cognitive load during decision-making allows individuals to reflect on the problem and deliberately select the correct and rational solution.	▪ Base rate neglect (1/1) ▪ Conjunction fallacy (1/1)	*Boissin et al. (2022):* Base rate neglect, Conjunction fallacy	
Repetition	Second intervention	▪ A second training further improves performance, by increasing learning input and fostering debiasing strategies.	▪ Base rate neglect (0/1) ▪ Conjunction fallacy (0/1)		*Boissin et al. (2022):* Base rate neglect, Conjunction fallacy

+ = Positive(ly moderating) effect on debiasing | o = no moderating effect on debiasing.
(a/b) = The positively moderating effect is found in a out of b studies.
[a] Studies that examine moderating factors are presented in this table. All other relevant studies are listed above

FIGURE 1
Mediators and moderators of training interventions on debiasing



<table>
<tr><td>Study no.</td><td colspan="4">4</td></tr>
<tr><td>Title</td><td colspan="4">Biases und Debiasing im M&A-Prozess - Die Identifikation von Biases und Debiasing-Maßnahmen im M&A-Prozess als Aufgabe des Controllings</td></tr>
<tr><td>Authors</td><td colspan="4">Marvin Göbel, Rebecca Sabel, Arnt Wöhrmann</td></tr>
<tr><td rowspan="7">Author contribution</td><td></td><td>Göbel</td><td>Sabel</td><td>Wöhrmann</td></tr>
<tr><td>Numeric share</td><td>0.3</td><td>0.6</td><td>0.1</td></tr>
<tr><td>Conceptual development of research question</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Theoretical integration</td><td></td><td>✓</td><td></td></tr>
<tr><td>Derivation of a normative model</td><td></td><td>✓</td><td></td></tr>
<tr><td>Elaboration of practical implications</td><td>✓</td><td>✓</td><td></td></tr>
<tr><td>Writing the manuscript</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Publication status</td><td colspan="4">Published
Göbel, Marvin/Sabel, Rebecca/Wöhrmann, Arnt (2022): Biases und Debiasing im M&A-Prozess. In: Controlling 34 (6), pp. 19–26 (VHB-Rating 2024: C)</td></tr>
<tr><td>Research approach</td><td colspan="4">Normative study</td></tr>
<tr><td>Language</td><td colspan="4">German</td></tr>
<tr><td>Abstract</td><td colspan="4">Entscheidungsträger unterliegen im M&A-Prozess zahlreichen Biases. Die Folge kann das Scheitern von Fusionen und Übernahmen sein. Der vorliegende Beitrag identifiziert relevante Biases, zeigt die phasenbedingte Anfälligkeit anhand einer Checkliste auf und diskutiert wirksame Debiasing-Maßnahmen mit dem Ziel der Rationalitätssicherung im M&A-Prozess als wichtige Aufgabe des Controllings.

English version
Decision-makers are subject to numerous biases in the M&A process. A consequence can be failing mergers and acquisitions. This paper identifies relevant biases, shows the phase-related vulnerability through a checklist, and discusses effective debiasing measures to ensure rationality in the M&A process as an essential task in management accounting.</td></tr>
</table>

Biases und Debiasing im M&A-Prozess

Die Identifikation von Biases und Debiasing-Maßnahmen im M&A-Prozess als Aufgabe des Controllings

Entscheidungsträger unterliegen im M&A-Prozess zahlreichen Biases. Die Folge kann das Scheitern von Fusionen und Übernahmen sein. Der vorliegende Beitrag identifiziert relevante Biases, zeigt die phasenbedingte Anfälligkeit anhand einer Checkliste auf und diskutiert wirksame Debiasing-Maßnahmen mit dem Ziel der Rationalitätssicherung im M&A-Prozess als wichtige Aufgabe des Controllings.

Marvin Göbel, Rebecca Sabel und Arnt Wöhrmann

Marvin Göbel, M. Sc., ist wissenschaftlicher Mitarbeiter an der Professur für Managerial Accounting an der *Justus-Liebig*-Universität Gießen.

Rebecca Sabel, M. Sc., ist wissenschaftliche Mitarbeiterin an der Professur für Managerial Accounting an der *Justus-Liebig*-Universität Gießen.

Prof. Dr. **Arnt Wöhrmann** ist Inhaber der Professur für Managerial Accounting an der *Justus-Liebig*-Universität Gießen.

1. Kognitive Verzerrungen als Ursache scheiternder M&A-Vorhaben

Fusionen und Übernahmen (im Folgenden: M&A-Transaktionen) sind ein breit genutztes Mittel des (vermeintlich) schnellen externen Unternehmenswachstums. Dabei steht der stetige Zuwachs an M&A-Transaktionen jedoch in Kontrast zur hohen Misserfolgsquote von ca. 50 % (vgl. *Schoenberg*, 2006, S. 366). *Mueller* (1997, S. 680) vertritt gar die Ansicht, dass die US-Wirtschaft genauso effizient (oder effizienter) wäre, wenn es in den letzten 50 Jahren (*Stand*, 1997) keine M&A-Transaktionen gegeben hätte. Daraus lässt sich schließen, dass sich die Erwartungen der Erwerber bei einer Vielzahl an Übernahmen nicht erfüllen. Ein möglicher Erklärungsansatz für das Scheitern zahlreicher Transaktionen ist die (unzutreffende) Unterstellung des Paradigmas des Homo Oeconomicus, nach dem Entscheider rational handeln und alle verfügbaren Informationen bei der Entscheidungsfindung berücksichtigen und richtig bewerten (vgl. *Beck*, 2014, S. 1 f.). Die Realität zeigt, dass Menschen regelmäßig irrational handeln, Entscheidungsheuristiken anwenden und gedankliche Abkürzungen nutzen. Sie unterliegen kognitiven Verzerrungen (Biases).

Der M&A-Prozess ist besonders anfällig für Biases. Beispielsweise kann das Scheitern einer Fusion durch exzessives Risikoverhalten aufgrund der Überschätzung der eigenen Fähigkeiten (Overconfidence Bias) begründet sein. In der Literatur werden weitere Erklärungsansätze für das Scheitern von M&A-Transaktionen angeführt, dabei wird jedoch den spezifischen Ursachen und den möglichen Gegenmaßnahmen wenig Beachtung geschenkt. Der vorliegende Beitrag möchte diese Lücke schließen. Zunächst wird der M&A-Prozess mit den jeweiligen Aufgaben kurz skizziert. Darauf aufbauend werden relevante Biases in den jeweiligen M&A-Phasen identifiziert und erläutert. Eine entwickelte Treiber-Checkliste soll Praktikern helfen, die Anfälligkeit für Biases anhand der Ausgestaltung der Projektorganisation zu erkennen. Zuletzt werden wirksame Debiasing-Maßnahmen vorgestellt.

2. Die M&A-Prozessphasen

Mergers and Acquisitions (M&A) können in drei Prozessphasen untergliedert werden (vgl. *Lucks/Meckl*, 2015, S. 98): (I) Konzeptphase, (II) Transaktionsphase und (III) Integrationsphase, die im Folgenden knapp skizziert werden.

Zur (I) Konzeptphase gehören die Teilprozesse der **Festlegung der Unternehmens- und Akquisitionsstrategie** sowie das **Screening**. So soll sichergestellt werden, dass das M&A-Vorhaben die Erreichung der Unternehmensziele unterstützt. Eine solche Unterstützung vorausgesetzt, werden die verfolgten Ziele (z. B. Internationalisierung) weiter spezifiziert und das Akquisitionsumfeld analysiert (vgl. *Eberl*, 2009, S. 81). Der Teilprozess **Screening** beschreibt die systematische Suche nach geeigneten Zielunternehmen.

Mit der Festlegung auf ein Zielunternehmen beginnt die (II) Transaktionsphase. Nach der ersten Kontaktaufnahme werden Vorverträge in Form eines „Letter of Intent“ eine einseitige Kaufinte-

Zentrale Aussagen

- Die hohe Misserfolgsquote von M&A-Transaktionen kann unter anderem aufgrund des Auftretens einer Vielzahl von Biases entlang des M&A-Prozesses erklärt werden.
- Die Anfälligkeit des Auftretens von Biases wird maßgeblich von (individuellen) Treibern und dem Fortschritt des M&A-Prozesses beeinflusst.
- Debiasing-Maßnahmen sind wichtige Controlling-Instrumente zur Rationalitätssicherung in der Entscheidungsfindung im Rahmen des M&A-Prozesses.

Motivational, Informational and Statistical Biases können in jeder Phase des M&A-Prozesses auftreten.

ressensbekundung – und eines „Memorandum of Understanding" – ein Vorvertrag über die Zwischenergebnisse der Vorverhandlungen – abgeschlossen. Anschließend werden die **Due Diligence** angestoßen und Informationen für die Prüfung des Zielunternehmens beschafft. Ziel ist es, Schwachstellen in der Planung zu erkennen. Darauf folgt die **Unternehmensbewertung** mithilfe eines geeigneten Bewertungsverfahrens (bspw. Discounted Cash-Flow-Verfahren). In diesem Zuge werden auch der Synergie- und Restrukturierungswert bestimmt. Der ermittelte Unternehmenswert dient dem Käufer als Referenzgröße für die **Verhandlungen**. Kommt es zu einer Preiseinigung, wird der Vertrag gezeichnet („Signing"), worauf kartell- und fusionsrechtliche Prüfungen folgen können. Die Transaktionsphase endet mit dem „Closing", d. h. der rechtlichen Wirksamkeit des Vertrags, und dem Übergang der Vertragsgewalt.

Die letzte Phase ist die (III) Integrationsphase, bei der zu Beginn die Aufbau- und Ablaufstruktur des neuen Unternehmens festgelegt und messbare Integrationsziele im Rahmen der **Integrationsplanung** bestimmt werden. Anschließend wird eine künftige Organisationsstruktur festgelegt, die, neben „harten Faktoren" (z. B. Personalfreisetzung), insbesondere die „weichen" Faktoren (z. B. kulturellen Wandel) regelt. Sodann erfolgen **Integrationsumsetzung** und **Integrationskontrolle**, d. h. die Prüfung der Zielerreichung.

3. Kognitive Verzerrungen entlang der M&A-Prozessphasen – Kategorisierung, Einordnung und Darstellung

Zur leichten Identifikation von Biases und Debiasing-Maßnahmen im M&A-Prozess werden Biases im Folgenden in **Motivational, Informational** und **Statistical Biases** eingeteilt (vgl. *Kreilkamp* et al., 2019, S. 57). Motivational Biases beschreiben (un)-bewusste Verzerrungen aufgrund von abweichenden Motiven der Entscheider, wie z. B. dem Verfolgen eigener Interessen. Informational Biases beeinflussen die Informationsauswahl und -verarbeitung, während Statistical Biases aufgrund von Fehleinschätzungen entgegen den Prinzipien der Wahrscheinlichkeitstheorie resultieren. Eine Analyse dieser Kategorien erscheint vor dem Hintergrund, dass einige Tätigkeiten im M&A-Prozess die gleichen Charakteristika aufweisen, sinnvoll. Daher sollen nicht die jeweiligen Phasen, sondern vielmehr die typischen Verzerrungen der einzelnen Aufgaben thematisiert werden (vgl. *Eberl*, 2009, S. 81). **Abb. 1** zeigt einen Überblick über die Phasen des M&A-Prozesses und das Auftreten von Biases, die im Folgenden diskutiert werden.

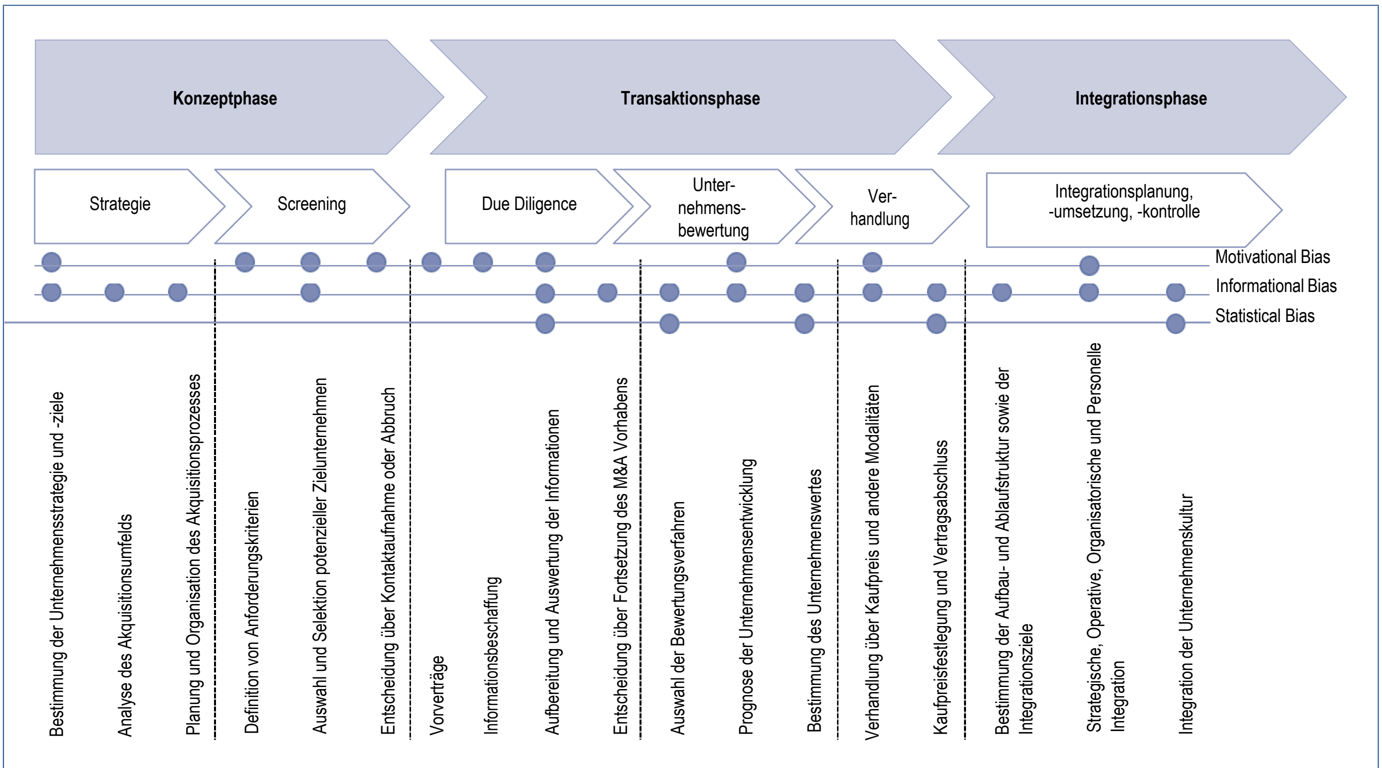


Abb. 1: Bias-Kategorien im M&A-Prozess (in Anlehnung an *Eberl*, 2009, S. 82)

Motivational Biases

Motivational Biases wie der **Confirmation Bias** lassen sich in jeder Phase finden. Im Rahmen der Unternehmens- und Akquisitionsstrategie wird analysiert, ob mit einem M&A-Vorhaben die Unternehmensziele erreicht werden könnten. Dabei besteht die Gefahr, dass Informationen im Sinne der Bestätigung der eigenen Meinung, z. B. getrieben durch vorangegangene Unternehmenskäufe, ausgewählt und gedeutet werden. Im Rahmen des Screenings wird die Suche nach geeigneten Zielunternehmen ggf. vorzeitig abgebrochen, wenn für den bestehenden Pool bereits erste positive Informationen vorliegen. Auch in der Transaktionsphase führt eine verkürzte und verzerrte Informationssuche dazu, dass fehlende Plausibilität und Risiken nicht aufgedeckt werden.

Entscheidungsträger präferieren M&A-Transaktionen auch deswegen, weil sie die eigene Macht und das Ansehen steigern (**Empire Building**). Externe Wachstumsoptionen sind jedoch im Vergleich zu den langsameren, internen Wachstumsoptionen oftmals riskanter. Hinzu kommt, insbesondere in der Integrationsphase, die Illusion, Umweltrisiken kontrollieren und abwenden zu können (**Illusion of control**). So scheitern z. B. viele M&As durch eine unzureichende kulturelle Integration (vgl. *Lodorfos/Boateng*, 2006, S. 1406).

Informational Biases

Ein Großteil der Informational Biases tritt in der Transaktions- und Integrationsphase auf. Zu den bekanntesten zählt der **Overconfidence** Bias, d. h. die systematische Überschätzung der eigenen Fähigkeiten. So wird in der Konzeptphase häufig die Wahrscheinlichkeit eines erfolgreichen Zusammenschlusses entgegen der empirischen Erfolgsquote zu optimistisch eingeschätzt oder in der Integrationsphase die Umsetzungsdauer unterschätzt (**Planning Fallacy**) (vgl. *Garbuio* et al., 2010, S. 92; *Eberl*, 2009, S. 84).

In der Due Diligence wird eine große Menge an Daten gesammelt und ausgewertet. Dabei kann das **Framing**, also die Art und Weise, wie Informationen dargestellt werden, die Aufbereitung und Verarbeitung der Informationen beeinflussen. Wird zum Beispiel der Einsatz von Mitarbeitertrainings im Rahmen der Integrationsphase diskutiert, fokussieren sich bspw. Manager primär auf den Kostenaspekt, während Personaler weiche Faktoren, bspw. die positive Motivationswirkung, in den Vordergrund stellen. Es besteht die Gefahr, dass eine Personengruppe dominiert, wodurch Informationen, den eigenen Wahrnehmungen entsprechend, verzerrt präsentiert werden.

Der **Anchoring and Adjustment Bias** ist insbesondere in der Transaktionsphase präsent. Dieser beschreibt, dass Individuen sich bei Schätzungen an einer irrelevanten Initialinformation (Anker) orientieren und diesen nur ungenügend anpassen. Der anfängliche Unternehmenswert wird häufig als Ankerwert verwendet, wodurch neue Informationen, die wertmindere Erkenntnisse offenlegen, ausgeblendet werden (vgl. *Asaoka*, 2019, S. 12).

Eine Unternehmenstransaktion ist mit Zeit, Arbeit und Kosten verbunden. Ein Einbeziehen dieser vergangenen Kosten bei Entscheidungen wäre irrational und wird als **Sunk Cost Fallacy** bezeichnet. Häufig werden nach der Konzeptphase die Kosten für die aufgewandten Ressourcen in den weiteren Schritten der Due Diligence unbemerkt miteinbezogen (vgl. *Garbuio* et al., 2010, S. 96). So wird die bereits aufgewandte Arbeitszeit als teure Investition wahrgenommen, weshalb ein Abbruch des M&A-Projektes als Verlust (in Höhe der Sunk Costs) missinterpretiert und gescheut wird. Stattdessen bieten Entscheidungsträger irrational hohe Preise für das Zielunternehmen. In diesem Zusammenhang wird auch vom **Winner's Curse** gesprochen. Dieser besagt, dass im Rahmen eines Wettbewerbs der Bieter den Zuschlag erhält, der den Wert des Zielunternehmens am stärksten überschätzt (vgl. *Varaiya/Ferris*, 1987, S. 65).

Ein nach oben verzerrter Unternehmenswert kann auf den Anchoring and Adjustment Bias und den Winner's Curse zurückzuführen sein.

Statistical Biases

Statistical Biases dominieren insbesondere die Transaktionsphase. Führungskräfte drohen bei der Due Diligence der **Availability Heuristic** zum Opfer zu fallen, wenn sie die Wahrscheinlichkeit eines Ereignisses alleine anhand der zur Verfügung stehenden Informationen beurteilen. So scheitern M&A-Prozesse oft an der Herausforderung, unterschiedliche (Unternehmens-) Kulturen zu vereinen (vgl. *Lodorfos/Boateng*, 2006, S. 1406). In der Due Diligence konzentrieren sich Führungskräfte bei der Analyse des M&A-Vorhabens primär auf Fragen der strategischen Vereinbarkeit. Jedoch führt der einseitige Fokus dazu, dass Fragen der organisatorischen Vereinbarkeit (bspw. der kulturellen Eignung) nicht hinreichend berücksichtigt werden, wodurch die Wahrscheinlichkeit einer erfolgreichen Integration überschätzt wird.

Ein weiterer Statistical Bias ist die **Overvaluation**. Um den Unternehmenswert und die zukünftigen Zahlungsüberschüsse ermitteln zu können, müssen die Bilanz- und GuV-Positionen prognostiziert werden. Aufgrund dieser präzisen Planung wird die Aussagekraft der eigenen Prognose überschätzt und die in der Planung immanente Unsicherheit übersehen. Häufig sind Manager davon überzeugt, bestehende Ineffizienzen aufdecken und neue Synergien erzeugen zu können. In Folge wird der Grenzpreis aus Käufersicht überschätzt, was ein Scheitern begünstigt.

4. Potenzielle Treiber aus Sicht der Projektaufbau- undAblauforganisation

Die Existenz und Auswirkungen der oben beschriebenen Biases sind für die Entscheidungsträger häu-

Die Gestaltung der Projektorganisation hat einen wesentlichen Einfluss auf das Auftreten von Biases im M&A-Prozess.

fig nicht klar erkennbar, weshalb es im Aufgabenbereich des Controllings liegt, ein generelles Bewusstsein für mögliche Rationalitätsdefizite im M&A-Prozess zu schaffen.

Mit Blick auf die bereits erfolgten Ausführungen erscheint der M&A-Prozess im Wesentlichen durch vier Haupttreiber anfällig für potenzielle Verzerrungen zu sein: (1) Beeinflussung durch individuelle Präferenzen, Motive und Überzeugungen (**M&A- Prozessstruktur; individuelle Entscheidungsmacht**); (2) strategische Relevanz des Vorhabens (**emotionales Involvement; individueller Zeit- und Kostendruck**); (3) Einbezug ausschließlich finanzwirtschaftlicher Kennzahlen (**M&A-spezifisches Informations- und Kommunikationssystem (IuK); kulturelle Vielfalt**); (4) Umgang mit Veränderungen und Widersprüchen (**organisatorische Resilienz**).

Im Folgenden werden die in Klammern genannten Einflussfaktoren und ihre Auswirkungen genauer erläutert und in eine Checkliste integriert. **Abb. 2** zeigt in welchem Ausmaß die einzelnen Treiber Motivational, Informational bzw. Statistical Biases befördern. Auf dieser Basis können Unternehmen die eigene Anfälligkeit abschätzen und sinnvolle Debiasing-Maßnahmen auswählen (vgl. **Abb. 3**).

Idealtypische Projektorganisationsformen sind gekennzeichnet durch interdisziplinäre Teamzusammensetzungen, Autonomiebestreben aller Beteiligten und rollenunabhängige, informelle Kommunikation (vgl. *Bea* et al., 2013, S. 9–11). Ist die **M&A-Prozessstruktur indes starr**, d. h. Entscheidungen werden von wenigen Entscheidern ohne ausreichendes Feedback in homogenen Teams getroffen, so ist denkbar, dass es z. B. in der Due Diligence (Transaktionsphase) zu einer einseitigen oder stark verkürzten Informationsauswahl und -verarbeitung kommt. Zusätzlich wird der Überzeugung einzelner zu große Bedeutung beigemessen und deren Sichtweise nicht hinreichend in Frage gestellt. In Verbindung mit aufbauorganisatorisch gesetzten Verantwortlichkeiten und Weisungsbefugnissen entsteht mitunter eine **hohe individuelle Entscheidungsmacht** einzelner Beteiligter, die insbesondere in der Konzeptphase dysfunktionale Handlungsspielräume erzeugen kann. In Folge entsteht Raum für nicht marktwertsteigernde, persönliche Machtmotive und deren Durchsetzung (Motivational Biases).

In der Transaktionsphase steigt mit Fortschreiten des M&A-Projekts und zunehmender Bindung finanzieller und personeller Mittel das **emotionale Involvement**. In Folge beschränkt sich die Informationssuche und Auswertung auf diejenigen Aspekte, die eine Fortführung des Projektes bestärken. Informationen, die die finanzielle Vorteilhaftigkeit in Frage stellen könnten, werden ausgeblendet (Informational Biases). Mit Blick auf das unternehmensexterne Handlungsfeld kann **wettbewerbsbedingter Zeit- und Kostendruck** als weiterer Treiber in der Transaktionsphase herausgestellt werden. Eine große Relevanz des Vorhabens im Unternehmen, begrenzte zeitliche, personelle und finanzielle Ressourcen bei zugleich geringer Anzahl an alternativen Zielunternehmen, können zu Handlungsdruck

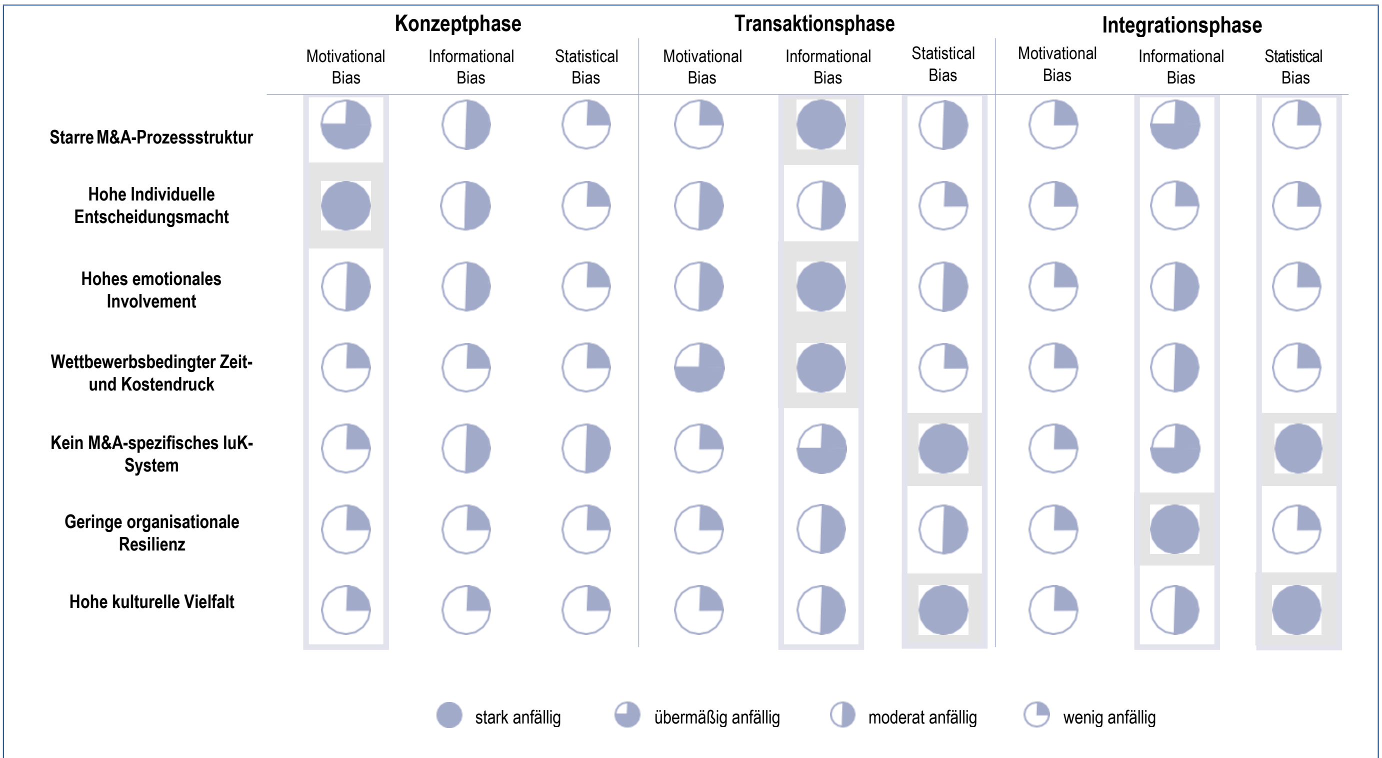


Abb. 2: Treiber zur Früherkennung von Bias-Risiken entlang der M&A-Prozessphasen

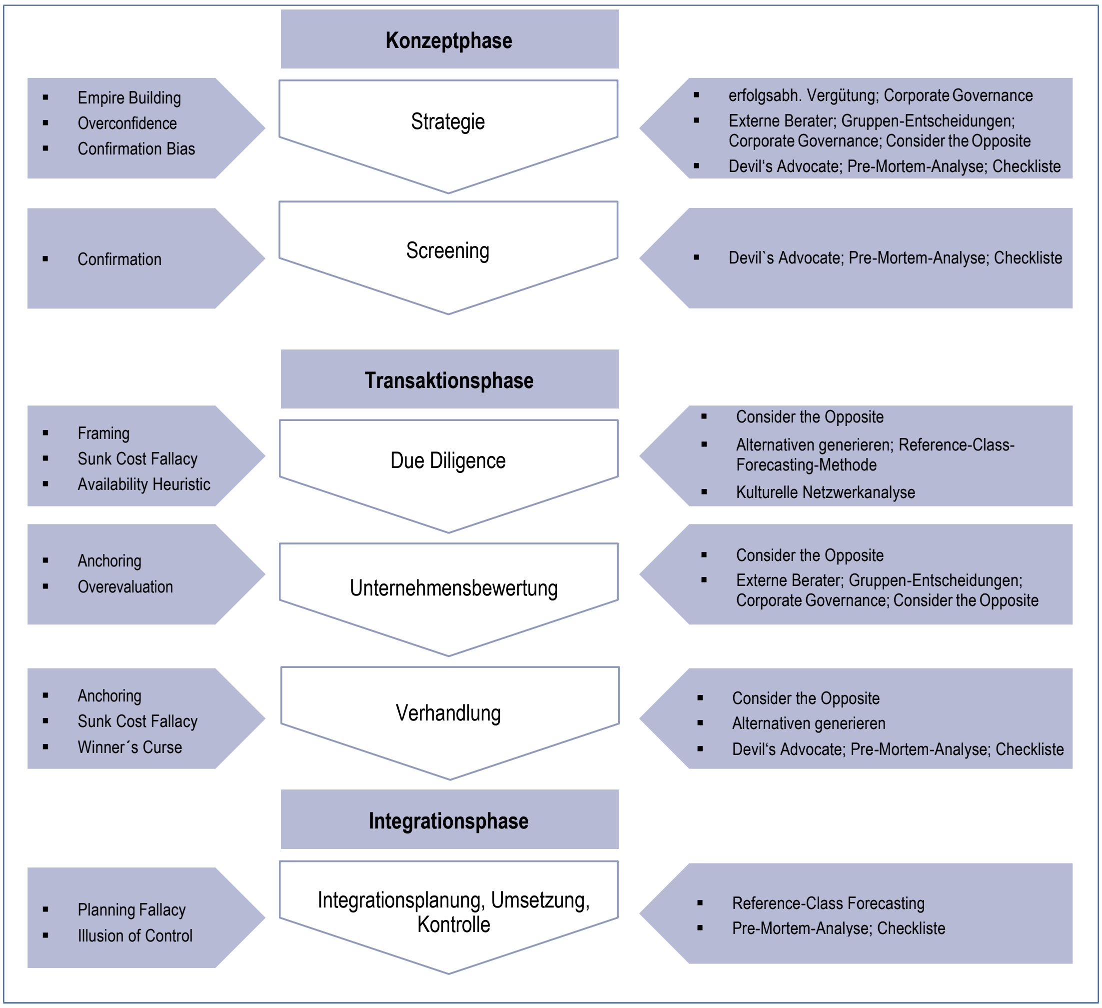


Abb. 3: Biases & Debiasing entlang des M&A-Prozesses

führen. Entscheidungen werden dann eher auf Grundlage motivationaler Aspekte und weniger durch eine umfängliche Betrachtung von Informationen und dem Einsatz mathematischer Auswertungsmethoden begründet. **IuK-Systeme** beeinflussen wie Informationen erfasst, weitergegeben und verarbeitet werden. Erfüllt ein solches System nicht die Anforderung eines neutralen und sensiblen Informationsaustauschs, besteht das Risiko einer einseitigen und unvollständigen Datenerfassung (Informational Biases) sowie -auswertung (Statistical Biases).

Im Rahmen der Integrationsphase erscheint der Umgang mit Veränderungen als ein wesentlicher Treiber für das Auftreten kognitiver Verzerrungen. Der organisationale Zusammenschluss zweier Unternehmen erfordert einen offenen Umgang mit auftretenden Herausforderungen bezogen auf die Wertschöpfungsprozesse, Kommunikationswege und -formen sowie bei internationalen Zusammenschlüssen auf die rechtlichen und politischen Besonderheiten. Eine **geringe organisationale Resili- enz**, also die fehlende Fähigkeit auf Veränderungen zu reagieren, kann die Erfolgschancen einer Inte- gration mindern (Informational Biases). Insbeson- dere hervorzuheben ist der Faktor der **kulturellen Vielfalt** zwischen Käufer- und Zielunternehmen. Werden wertvolle Ressourcen im Falle einer hohen kulturellen Heterogenität ausschließlich für finanzwirtschaftliche Aspekte und somit zu wenig für die interkulturelle Kommunikation verwendet (Informational und Statistical Biases), droht das Vorhaben auf der Zielgeraden zu scheitern.

Zusammenfassend wird deutlich, dass die Konzeptphase Risiken für Motivational Biases birgt, während mit zunehmendem Projektverlauf in der Transaktions- und Integrationsphase Informational und Statistical Biases dominieren (vgl. Hervorhebungen der Kreise in **Abb. 2**).

5. Rationalitätssicherungdurch Debiasing – eine originäre Aufgabe des Controllings

Die beschriebenen Biases und die resultierenden Rationalitätsdefizite belegen die Notwendigkeit für Gegenmaßnahmen. Dabei rückt die originäre Aufgabe des Controllings bei der Rationalitätssicherung der Führung in den Fokus. Neben der Entscheidungsunterstützung ist eine Verhaltenssteuerung erforderlich. Hierfür kann das Controlling auf das Management Control System, also das Steue-

Neben der Entscheidungsunterstützung ist das Controlling für die Verhaltenssteuerung verantwortlich.

Das Controlling kann durch den Einsatz von Kontrollinstrumenten und einer effizienten Anreizsteuerung kognitiven Verzerrungen entgegenwirken.

rungssystem eines Unternehmens, das genutzt wird, um die Unternehmensziele zu erreichen, zurückgreifen (vgl. *Merchant/Van der Stede*, 2017). Analog zu den klassischen Controlling-Instrumenten sollen hierbei insbesondere verhaltensorientierte Mechanismen in Form von Debiasing-Methoden angewandt werden. Im M&A-Kontext setzen diese entweder über **Anreize** an den Motiven des Einzelnen an, um diese in Einklang mit den Unternehmenszielen zu bringen, oder beeinflussen mittelbar das Verhalten der Mitarbeiter durch **Kontrollinstrumente**.

Neben regelhaften Maßnahmen in Form von Corporate Governance oder Rechtfertigung fördern die Instrumente eine **analytisch-reflektierende Denkweise** (allgemeine **Aufmerksamkeitssteuerung** und **Einbezug Dritter**). Im Folgenden werden exemplarisch Debiasing-Maßnahmen vorgestellt. **Abb. 3** gibt einen Überblick über die identifizierten Biases im M&A-Prozess und die jeweils geeigneten Debiasing-Instrumente.

Motivational Debiasing

Monetäre Anreize, insbesondere in Form erfolgsabhängiger Vergütungsbestandteile, stellen ein traditionelles Controlling-Instrument der Anreizsteuerung dar. Durch Bindung der Vergütung an den Erfolg von M&A-Transaktionen werden in der Konzeptphase nicht-marktwertsteigernde Motive zurückgedrängt, die auf inneren Beweggründen wie Macht und Prestige beruhen (Empire-Building). Darüber hinaus kann im Rahmen der **Corporate Governance** ein Abstimmungsprozess mit dem Aufsichtsrat oder eine Genehmigungspflicht durch das Zentralcontrolling implementiert werden. Zum einen wird der Manager hierdurch verpflichtet, die unternehmenswertsteigernden Beweggründe seiner Entscheidung vorzubringen (**Rechtfertigungskontrolle**). Hierbei ist das Controlling für die Prämissenkontrolle zuständig. Zum anderen wird eine konstruktive Diskussion des Vorhabens mit Blick auf die Bedürfnisse der unterschiedlichen Anspruchsgruppen des Unternehmens ermöglicht (vgl. *Garbuio* et al., 2010, S. 87 f.).

Insbesondere in der Transaktionsphase besteht die Gefahr, dass trotz neuer Informationen im Rahmen der Due Diligence an ursprünglichen Einschätzungen hinsichtlich des Kaufpreises (Confirmation Bias) oder gar gesamten, nunmehr nicht erfolgsversprechenden, M&A-Transaktion festgehalten wird (Winner's Curse). Eine mögliche Maßnahme zur Reduzierung dieser Biases ist die sogenannte **Devil's Advocate-Methode** (vgl. *Beckhaus*, 2013, S. 207 f.). Dabei argumentiert eine nicht direkt von den Folgen des M&A-Prozesses betroffene Person, bspw. eine Controllerin, bewusst gegen das M&A-Vorhaben, um Gefahren herauszustellen. Ähnlich wirkt die **Pre-Mortem-Analyse** (vgl. *Klein*, 2007, S. 19), bei der im Rahmen eines Gedankenexperiments die Annahme gesetzt wird, dass das M&A-Projekt zukünftig scheitert. Anschließend werden mögliche Gründe für das Scheitern identifiziert. Dies führt zu einer besseren Abwägung. Von dem Akquisitionscontrolling vorab definierte **Checklisten** (vgl. *Beckhaus*, 2013, S. 207) können zudem dabei unterstützen, sämtliche kaufpreisrelevanten Informationen einzubeziehen. Die beiden letztgenannten Debiasing-Maßnahmen erscheinen auch geeignet, kulturrelevante Aspekte einzubinden und somit eine in der Integrationsphase auftretende Kontrollillusion (Illusion of Control) zu reduzieren.

Informational Debiasing

Informational Biases werden durch verzerrte Informationsselektion und Denkprozesse begünstigt, die durch das Handlungs- und Entscheidungsumfeld geprägt sind. Im Rahmen der Informationsfunktion ist es die Aufgabe des Controllings Informationen zu sammeln und unverzerrt aufzuarbeiten. Mit Blick auf das Debiasing ist es erforderlich, die Aufmerksamkeit des Entscheiders weg von persönlichen Einstellungen und hin zu einer sachlich-argumentativen Denkweise zu bewegen. Als geeignet erscheinen auch hier die obigen Instrumente. So ist es denkbar, dass dem Overconfidence Bias dadurch begegnet werden kann, dass Dritte in Form von **externen Beratern**, **Gruppen-Entscheidungen** oder über die **Corporate Governance** in den Entscheidungsprozess eingebunden werden und z. B. Synergien beurteilen. Gegenteilige Argumente (**Consider the Opposite**) (vgl. *Kreilkamp* et al., 2019, S. 61) oder Worst Case-Szenarien (**Pre-Mortem Analyse**) können auf individueller Ebene eingesetzt werden, um einen Erfolgstunnelblick zu mindern und Risiken in Bezug auf Synergiepotenziale bewusster zu machen. Ersteres erscheint auch geeignet, um Framing oder Anchoring and Adjustment bei der Kaufpreisbestimmung zu begegnen. Das Controlling hat hier die Aufgabe, die unterschiedlichen Wahrnehmungsrahmen zu kompensieren und relevante Werttreiber, bspw. die Übernahmekosten, in den Fokus zu rücken, um Ankereffekte zu vermeiden (vgl. *Eberl*, 2009, S. 86).

Darüber hinaus sollten im Auswahlprozess weitere potenzielle Zielunternehmen nicht zu früh verworfen werden, um **Alternativen zu generieren**. Eine breitere Vergleichsbasis und eine unabhängigere Verhandlungsposition durch geringeren Kaufdruck sind die Folge. Des Weiteren wird die Wahrscheinlichkeit einer emotionalen Bindung reduziert, sodass Sunk Costs eine geringere Rolle spielen (vgl. *Garbuio* et al., 2010, S. 96 f.). Eine Möglichkeit des Außenvergleichs bildet das **Reference-Class-Forecasting** (vgl. *Lovallo/Kahneman*, 2003, S. 61 f.). So ist es denkbar, dass Zielunternehmen anhand vorab definierter Parameter (bspw. Zeit und Kosten) mit bereits abgeschlossenen M&A-Transaktionen als Referenzklasse verglichen werden. Dies gewährleistet realistische Prognosen, was genauere Zeit- und Kostenpläne ermöglicht.

Statistical Debiasing

Statistische Verzerrungen münden in fehlerhafte Berechnungen der Synergiepotenziale und somit des Unternehmenswerts. Eines der wesentlichen Probleme ist, dass weiche Faktoren (wie z. B. die Prüfung des Kultur-Fits) aufgrund ihres qualitativen Charakters unzureichend in der M&A-Mehrwertbestimmung berücksichtigt werden. Die Durchführung einer **kulturellen Due Diligence** in Form von **Netzwerkanalysen** kann dieses Problem adressieren. Dabei werden die organisationalen Verbindungen der Mitarbeiter untersucht. Dies liefert Aufschluss über das Ausmaß kultureller Ähnlichkeiten, die vorherrschende Mitarbeiterbindung und bestehende Netzwerke. Da diese Informationen ex-ante häufig nicht verfügbar sind, beschränkt sich die Analyse mittelbar auf frei zugängliche Veröffentlichungen (vgl. *Garbuio* et al., 2010, S. 94). Hier ist es vorteilhaft, wenn das Controlling im Kontakt mit den im M&A-Prozess eingebundenen Personen steht, um auch solche weichen Faktoren zu berücksichtigen.

6. Abschließende Betrachtung

Die Anzahl von Übernahmen und Fusionen wächst stetig. Neben der Zielsetzung der Verbesserung der eigenen Marktposition ist der Prozess durch menschliches Verhalten geprägt. Entscheidungsträger unterliegen begrenzter Rationalität. Individuelle Präferenzen und Motive sowie beschränkte Kapazitäten der kognitiven Informationsverarbeitung sind mögliche Quellen kognitiver Verzerrungen (Biases), die auch vor dem M&A-Prozess nicht Halt machen. Es ist die Aufgabe des Controllings einen rationalen Ablauf sicherzustellen.

Der Beitrag schafft, basierend auf der Darstellung eines dreistufigen Prozessmodells, ein Verständnis potenziell relevanter, kognitiver Verzerrungen. Diese haben ihren Ursprung entweder in nicht wertsteigernden, individuellen Präferenzen des Entscheiders (Motivational Biases), sind durch eine Verzerrung der Informationsauswahl und -verarbeitung begründet (Informational Biases) oder beruhen auf der mangelhaften Anwendung statistischer Methoden (Statistical Biases). Mit Blick auf die M&A-Projektorganisation wurden Treiber herausgearbeitet, die an der Aufbauorganisation (Hierarchie, Verantwortung) und der Ablauforganisation (Informationsflüsse, Informationsverarbeitung) ansetzen. Anschließend wurde eine Checkliste vorgestellt, mit der Unternehmen prüfen können, welche M&A-Teilprozesse besonders anfällig für Verzerrungen sind. Im letzten Schritt wurden exemplarisch Gegensteuerungsmaßnahmen vorgestellt, die entweder im Sinne traditioneller Anreiz- und Kontrollmechanismen (z. B. Corporate Governance, erfolgsabhängige Vergütung), im sozialen Austausch mit Dritten (z. B. externe Berater, Gruppenentscheidungen) oder durch Aufmerksamkeitssteuerung auf Ebene des einzelnen Entscheiders (z. B. Consider the Opposite, Alternativen schaffen) ansetzen. Hier liegt es in der Verantwortung des Akquisitionscontrollings, neben dem Schaffen des Bewusstseins für Biases, die Rationalität des M&A-Prozesses zu gewährleisten.

Es wurde gezeigt, dass insbesondere zu Beginn des M&A-Prozesses motivationale Biases auftreten, die ihren Ursprung in hierarchischen Entscheidungsbefugnissen haben. Traditionelle monetäre Anreizinstrumente können die Kompatibilität wiederherstellen. Mit fortschreitendem Projektverlauf und zunehmender Konkretisierung rücken Informational Biases in den Fokus. Auf ablauforganisatorischer Ebene besteht die Gefahr darin, dass bewertungsrelevante Informationen gar nicht oder durch zu starke Standardisierung zu eng ausgewählt und ausgewertet werden (Statistical Biases). Hier scheinen insbesondere diejenigen Maßnahmen geeignet, die die Bildung einer objektiven, professionellen Sichtweise sowie den Einsatz analytischer Ansätze begünstigen.

Implikationen für die Praxis

- Unternehmen sollten in jeder Phase des M&A-Prozesses untersuchen, ob und welche Biases vorliegen könnten. Dabei kann die dargestellte Checkliste helfen.
- Debiasing-Maßnahmen ermöglichen die Sicherung der Rationalität bei der Entscheidungsfindung, wodurch effiziente Entscheidungen getroffen und Kosten eingespart werden können.
- Die Verteilung der Ressourcen für den Einsatz von Debiasing Maßnahmen sollte kontinuierlich entlang dem M&A-Prozess basierend auf unternehmensindividuellen Treibern und dem M&A-Fortschritt angepasst werden.

Die ungenügende Beachtung „weicher Faktoren", wie beispielsweise der kulturellen Passung, führt häufig zum Scheitern von M&A-Transaktionen.

Literatur

- *Asaoka, D.*, Behavioral analysis of mergers and acquisitions decisions, in: Corporate Board role duties and composition (CB), 15. Jg. (2019), H. 3, S. 8–16.
- *Bea, F.X./Scheurer, S./Hesselmann, S.*, Praxis der Projektorganisation: Projektmanagement konkret, 1. Aufl., Konstanz, München 2013.
- *Beck, H.*, Behavioral Economics. Eine Einführung, Wiesbaden 2014.
- *Beckhaus, G.*, Comply or Explain – A Flexible Mechanism to Countervail Behavioral Biases in M&A Transactions, in: University of Miami Business Law Review, 21. Jg. (2013), H. 1, S. 183–221.
- *Eberl, S.*, Der M&A-Prozess: Struktur, Rationalitätsdefizite und die Rolle des Controllings, in: Zeitschrift für Controlling und Management, 53. Jg. (2009), H. 1, S. 80–87.
- *Garbuio, M./Lovallo, D./Horn, J.*, Overcoming biases in M&A: A process perspective, in: Advances in Mergers and Acquisitions, 9. Jg. (2010), S. 83–104.

- *Klein, G.*, Performing a Project Premortem, in: Harvard Business Review, 95. Jg. (2007), H 9, S. 18–19.
- *Kreilkamp, N./Schmidt, M./Wöhrmann, A.*, Effizienzsteigerung durch Debiasing: Empfehlungen für die Praxis, in: Controlling, 31. Jg. (2019), H. 2, S. 57–65.
- *Lodorfos, G./Boateng, A.*, The role of culture in the merger and acquisition process, in: Management Decision, 44. Jg. (2006), H. 10, S. 1405–1421.
- *Lovallo, D./Kahneman, D.*, Delusions of Success: How Optimism Undermines Executives Decisions, in: Harvard Business Review, 81. Jg. (2003), H. 7, S. 56–63.
- *Lucks, K./Meckl, R.*, Internationale Mergers & Acquisitions. Der prozessorientierte Ansatz, 2. Aufl., Berlin, Heidelberg 2015.
- *Merchant, K. A./Van der Stede, W. A.*, Management Control Systems, 4. Aufl. Harlow 2017.
- *Mueller, D. C.*, Merger Policy in the United States: A Reconsideration, in: Review of Industrial Organization, 12. Jg. (1997), H. 5/6, S. 655–685.
- *Schoenberg, R.*, Measuring the Performance of Corporate Acquisitions: An Empirical Comparison of Alternative Metrics, in: British Journal of Management, 17. Jg. (2006), H. 4, S. 361–370.
- *Varaiya, N. P./Ferris, K. R.*, Overpaying in Corporate Takeovers: The Winner's Curse, in: Financial Analysts Journal 43. Jg. (1987), H. 3, S. 64–70.

Literaturtipps aus dem Online-Archiv
http://elibrary.vahlen.de

- *Niklas Kreilkamp, Maximilian Schmidt, Arnt Wöhrmann,* Effizienzsteigerung durch Debiasing: Empfehlungen für die Praxis, Ausgabe 2/2019, S. 57–65
- *Hanno Drews, Morten Friedrichsen,* Budgetierung und selektive Wahrnehmung, Ausgabe 2/2006, S. 79–84
- *Andreas Aschenbrücker,* Behavioral Controlling, Ausgabe 3/2012, S. 192–193

Stichwörter

Debiasing # Fusionen und Übernahmen # kognitive Verhaltensverzerrungen # M&A # Rationalitätssicherung

Keywords

biases # debiasing # M&A # mergers and acquisitions # rationality assurance

Summary

Decision-makers are subject to numerous biases in the M&A process. A consequence can be failing mergers and acquisitions. This paper identifies relevant biases, shows the phase-related vulnerability through a checklist, and discusses effective debiasing measures to ensure rationality in the M&A process as an essential task in management accounting.

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.

Signature author (Rebecca Sabel) Date

Submitted Papers:

1. Sabel, Rebecca/Wöhrmann, Arnt/Gerstel, Hannes (2023): *Celebrating Failure – The Effects of Failure Awards on Risk-Taking and Escalation of Commitment.* Working Paper (R&R in Contemporary Accounting Research).

2. Sabel, Rebecca/Kahl, Niklas/Wöhrmann, Arnt/Ewelt-Knauer, Corinna (2025): *The Effect of Monitoring on Teleworkers' and Office Workers' Behavior*. Working Paper (Submitted to Journal of Management Accounting Research).

3. Sabel, Rebecca/Wöhrmann, Arnt/Alt, Marlene (2025): *Debiasing Training—Mitigating Cognitive Biases Via Training Interventions: A Systematic Literature Review.* Working Paper (R&R in Accounting Perspectives).

4. Göbel, Marvin/Sabel, Rebecca/Wöhrmann, Arnt (2022): *Biases und Debiasing im M&A-Prozess*. Controlling 34 (6), pp. 19–26.