

**Taking a second look – Revising the effects of process and  
outcome accountability**

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

–

im Promotionsprogramm Behavior and Cognition

der Georg-August University School of Science (GAUSS)

vorgelegt von

Johannes Rollwage

aus Salzgitter

Göttingen, 2022

## **Betreuungsausschuss**

Prof. Dr. Stefan Schulz-Hardt, Abteilung Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. Andreas Mojzisch, Institut für Psychologie, Universität Hildesheim

Dr. Dietmar Zinner, Kognitive Ethologie, Deutsches Primatenzentrum

## **Mitglieder der Prüfungskommission**

Referent: Prof. Dr. Stefan Schulz-Hardt, Abteilung Wirtschafts- und Sozialpsychologie, Georg-Elias-Müller-Institut für Psychologie

Korreferent: Prof. Dr. Andreas Mojzisch, Institut für Psychologie, Universität Hildesheim

## **Weitere Mitglieder der Prüfungskommission:**

Dr. Dietmar Zinner, Kognitive Ethologie, Deutsches Primatenzentrum

Prof. Dr. Margarete Boos, Abteilung Sozial- und Kommunikationspsychologie, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. Lars Penke, Abteilung Biologische Persönlichkeitspsychologie, Georg-Elias-Müller-Institut für Psychologie

Prof. Dr. York Hagmayer, Abteilung Kognitionswissenschaft und Entscheidungspsychologie, Georg-Elias-Müller-Institut für Psychologie

Tag der mündlichen Prüfung: 05.12.2022

## **Danksagung**

„Rom wurde nicht an einem Tag erbaut“, so ähnlich ist es auch mit meiner Dissertation, auch die hat so ihre Zeit gebraucht um nun endlich vollendet zu sein. Aber viel wichtiger noch, Rom wurde nicht von einem Einzelnen erbaut, genau so wenig kann man eine bzw. diese Dissertation als das Werk eines Einzelnen verstehen. Denn auch wenn sie am Ende mit meinem Namen versehen ist, so ist sie doch nur deshalb möglich, weil eine Vielzahl von Menschen mich auf dem Weg und bei der Fertigstellung dieser Arbeit begleitet und unterstützt haben. All diesen Menschen möchte ich an dieser Stelle meine tiefe Dankbarkeit ausdrücken.

Zunächst gilt mein besonderer Dank selbstverständlich meinem Doktorvater Stefan Schulz-Hardt, der mir nicht nur das Vertrauen geschenkt hat als „Fachfremder“ in seiner Abteilung zu promovieren, sondern der mir im Rahmen meiner Promotionen stets mit unerlässlichen Anregungen, Ratschlägen und manchmal auch wichtiger Kritik zur Seite stand. Die Zusammenarbeit mit ihm hat mich in vielerlei Hinsicht geprägt und mein Selbstvertrauen sowie auch Selbstverständnis als Wissenschaftler gestärkt.

Des Weiteren danke ich auch den Mitgliedern meines Thesis Committees, Andreas Mojzisch und Dietmar Zinner, für ihr anhaltendes Interesse an meinem Forschungsprojekt, sowie für ihre bereichernden Anmerkungen und ihre generelle Unterstützung.

Margarte Boos, Lars Penke und York Hagemeyer danke ich für ihre Bereitschaft, Teil meines Prüfungskomitees zu sein.

Meinem Kollegen Thomas Schultze-Gerlach danke ich ganz besonders dafür, dass er mein Interesse an der Wissenschaft neu entfacht hat und mir den Weg für den Einstieg in die psychologische Forschung geebnet hat. Zusätzlich danke ich ihm für die Unterstützung in den letzten Jahren gerade auch in Form von vielen wichtigen Denkanstößen.

Ganz spezieller Dank gebührt meinem Freund, Kollegen, und manchmal auch menschengewordenen Kummerkasten, Christian Treffenstädt der immer ein offenes Ohr für meine Probleme hatte und mich stets mit gutem Rat und noch besserer Tat unterstützt hat. Und dessen stets einsatzbereiter Kaffeevollautomat mir über so manches Nachmittagstief beim Schreiben dieser Dissertation hinweggeholfen hat.

Danken möchte ich aber auch meinen Freunden Addi, Christoph, Jan und ganz besonders Bene und Irina, dafür dass ihre Türen mir immer offen standen, wenn ich Ablenkung und Aufheiterung brauchte oder auch einfach mal meinen Frust ablassen und ertränken musste.

Weiterhin danke ich meinen Schwiegereltern Lothar und Dorothea deren Unterstützung mir zeitlich wichtige Freiräume gegeben hat, die für die Anfertigung dieser Dissertation unerlässlich waren.

Auch meinen eigenen Eltern möchte ich selbstverständlich von ganzem Herzen danken. Ohne ihre, sicherlich nicht immer leichte, Erziehungsarbeit und ihre stetige Unterstützung in allen Lebenslagen hätte diese Dissertation niemals das Licht der Welt erblicken können und ich wäre nicht zu dem Menschen geworden, der ich heute bin.

Last but **definitely not** least, möchte ich meiner Frau Melanie und meinem Sohn Bjarne danken. Ihr habt, egal wie sehr ich auch an mir und meiner Arbeit gezweifelt habe, immer an mich geglaubt und mir euer tiefstes und ehrlichstes Vertrauen geschenkt. Ihr habt mich, wenn ich am Boden lag wieder aufgebaut. Nur durch euch war ich in der Lage diesen Weg zu bestreiten. Ihr habt mir die nötige Kraft gegeben auch in den dunkelsten Augenblicken der letzten Jahre weiter zu machen, mich durchzubeißen und nach vorne zu blicken. Eure Umarmungen, euer Lachen und eure Liebe, wenn ich nach einem harten Tag am Schreibtisch nach Haus komme erinnert mich stets daran wofür ich das hier tue.

Danke!

Göttingen, im November 2022

Johannes Rollwage

## Table of Contents

<b>Preface</b> .....	2
<b>Summary</b> .....	3
<b>Deutsche Zusammenfassung</b> .....	4
<b>1. Introduction</b> .....	6
1.1 Accountability .....	7
1.2 Process and outcome accountability .....	10
1.3 The problem with accountability research.....	12
<b>2. Summary of Manuscript 1: “Does holding decision-makers accountable make them less loss averse? A three-study replication attempt.”</b> .....	16
<b>3. Summary of Manuscript 2: “What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms.”</b> .....	19
<b>4. General Discussion</b> .....	23
4.1 Implications .....	24
4.2 Limitations and Future Research .....	29
<b>5. Conclusion</b> .....	33
<b>6. References</b> .....	34

**Appendix A.** Manuscript 1 (Does holding decision-makers accountable make them less loss averse? A three-study replication attempt.)

**Appendix B.** Manuscript 2 (What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms.)

**Appendix C.** Curriculum Vitae

## Preface

This dissertation follows the form of a cumulative dissertation. It includes two manuscripts, the first of which has been submitted for publication. The following manuscripts are included in Appendices A and B:

Rollwage, J., Häusser, J. A., & Schulz-Hardt, S. (2022). *Does holding people accountable make them less loss averse? A three-study replication attempt*. [Manuscript submitted for publication], Department of Economic and Social Psychology, University of Goettingen.

Rollwage, J., Treffenstädt, C., & Schulz-Hardt, S. (2022). *What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms*. [Unpublished manuscript], Department of Economic and Social Psychology, University of Goettingen.

In this dissertation, I integrate the theoretical background of these manuscripts and discuss their broader implications for the research on the effects of process and outcome accountability. I am the first author of both of these manuscripts.

For the three experiments in the first manuscript, I collected the data for the third experiment, provided the data analysis for all experiments, and wrote the manuscript. My co-authors contributed the study designs, collected the data of the first two experiments, and provided essential feedback during the entire project, especially during the writing process. For the second manuscript, I designed the study, collected and analyzed the data, and wrote the manuscript. My co-authors provided extensive and important feedback during all of these stages.

All parts of this dissertation were written by myself, the assistance of third parties was only accepted if scientifically justifiable and acceptable with regard to the examination regulations, and all sources have been quoted.

## Summary

Accountability is an essential aspect of human judgment and decision-making. Past research has claimed that different forms of accountability have differential effects on the quality of the decisions people make. Specifically, it has been reported that holding people accountable for their decision-making process (process accountability) rather than the outcomes of their decisions (outcome accountability) leads to higher decision quality and reduces cognitive biases. However, recently this claim, which has found entry into multiple influential literature reviews and textbooks, has been called into question by a series of unsuccessful replication attempts. This dissertation provides further evidence that previous findings regarding the beneficial effects of process accountability need to be scrutinized and tested. As in Manuscript 1, my co-authors and I found evidence that a classic finding from the accountability literature, Vieider's (2009) finding that accountability can reduce loss aversion, does not replicate. In Manuscript 2, we also discussed and tested the potential shortcomings of one of the most common experimental paradigms in accountability research that all of the studies that have failed to replicate so far share. In this paradigm, process and outcome accountability are induced by announcing interviews to the study participants. In the two experiments reported in the second manuscript, we found clear evidence that this "interview manipulation" of process and outcome accountability does not reliably produce the psychological effects it is supposed to induce. Therefore, the results reported in this dissertation impose serious limitations on the generalizability of previous findings regarding the effects of process and outcome accountability. While more theoretical and empirical work is needed to clarify how to address these problems regarding the replicability of previous findings of the effects of process and outcome accountability and if there are any more problems with this research, this work identifies important methodological problems within accountability research and discusses potential future directions to solve these problems.

## Deutsche Zusammenfassung

Verantwortlichkeit ist ein wesentlicher Bestandteil der menschlichen Urteils- und Entscheidungsfindung. In der bisherigen Forschung zu Verantwortlichkeitseffekten wurde die Behauptung aufgestellt, dass verschiedene Formen von Verantwortlichkeit unterschiedliche Auswirkungen auf die Qualität der von Menschen getroffenen Entscheidungen haben. Genauer wird in der Literatur berichtet, dass Verantwortlichkeit für den Entscheidungsprozess (Prozessverantwortlichkeit) im Gegensatz zu Verantwortlichkeit für die Ergebnisse der Entscheidungen (Ergebnisverantwortlichkeit) zu einer höheren Entscheidungsqualität führt und kognitive Verzerrungen verringert. In jüngster Zeit wurde diese Behauptung, die Eingang in zahlreiche einflussreiche Literaturreviews und Lehrbücher gefunden hat, jedoch durch eine Reihe erfolgloser Replikationsversuche in Frage gestellt. Diese Dissertation liefert weitere Beweise dafür, dass frühere Befunde bezüglich der positiven Auswirkungen von Prozessverantwortlichkeit stärker hinterfragt und getestet werden müssen. Meine Co-Autoren und ich finden in Manuskript 1 dieser Dissertation Belege dafür, dass ein klassischer Befund aus der Verantwortlichkeits-Literatur, Vieiders (2009) Befund, dass Personen, wenn sie sich für ihre Entscheidungen verantworten müssen weniger Verlustaversion zeigen, nicht repliziert werden kann. Darüber hinaus haben wir in Manuskript 2 einige potenzielle Mängel eines der gängigsten experimentellen Paradigmen in der Verantwortlichkeitsforschung diskutiert und getestet. Dieses Paradigma, welches in allen bisher nicht replizierbaren Studien verwendet wurde, induziert Prozess- und Ergebnisverantwortlichkeit durch die Ankündigung an die Studienteilnehmer, dass mit ihnen ein Interview geführt werden soll in dem sie sich für die Ergebnisse ihrer Entscheidungen bzw. den Prozess hinter ihren Entscheidungen rechtfertigen müssen. In den beiden Experimenten, über die im zweiten Manuskript berichtet wird, fanden wir eindeutige Belege dafür, dass diese "Interview-Manipulation" von Prozess- und Ergebnisverantwortung nicht zuverlässig die psychologischen Effekte hervorruft, die sie



eigentlich hervorrufen soll. Die in dieser Dissertation berichteten Ergebnisse schränken daher die Generalisierbarkeit bisheriger Erkenntnisse über die Auswirkungen von Prozess- und Ergebnisverantwortung erheblich ein. Während weitere theoretische und empirische Arbeiten erforderlich sind, um zu klären, wie diese Probleme bezüglich der Reproduzierbarkeit früherer Erkenntnisse über die Auswirkungen von Prozess- und Ergebnisverantwortlichkeit adressiert werden können und ob es noch weitere Probleme im Zusammenhang mit dieser Forschung gibt, zeigt diese Arbeit dennoch wichtige methodische Probleme innerhalb der Verantwortlichkeitsforschung auf und identifiziert mögliche Lösungen.

## 1. Introduction

*“[...] replication sit[s] at the heart of scientific progress [...]”*

(Walker et al., 2017)

The endeavor to replicate and try to disprove earlier work is central to scientific discovery and progress. However, as has been extensively discussed in recent years, it seems that large parts of the scientific community, especially certain fields, have lost sight of this central component of scientific progress. Fields such as medicine, economics, and, most prominently, psychology have seen a "replication crisis" (Pashler & Wagenmakers, 2012). This crisis has many facets, but in its most basic form, it describes the realization that a substantial part of the research reported in psychology (and, for what it is worth, in most other scientific fields) might be false (Ionnadis, 2005) and, thus, not replicable. Therefore, it seems essential for researchers in psychology to question earlier findings and try to replicate them to learn which parts of the psychological literature are replicable and, thus, reliable. However, replication studies still make up only a minuscule part (about 1%) of the psychological literature (Makel et al., 2012), and there are still numerous fields in psychology that almost exclusively consist of original findings and lack direct replication attempts. One prominent field in psychology that, for the longest time, had very much fit this description and has recently seen a number of unsuccessful replication attempts is the research of accountability. This dissertation aims to provide further evidence regarding the robustness and replicability of findings from the accountability literature, specifically regarding the robustness of the effects of process vs. outcome accountability. Furthermore, it identifies and tests methodological problems in the research of process and outcome accountability that could be responsible for the difficulties researchers seem to have in replicating the findings of earlier accountability research.

## 1.1 Accountability

Accountability is a concept that has been investigated since at least the late 70s of the previous century and nowadays is somewhat of a buzzword. While there are a couple of different definitions and understandings of the concept of accountability, at least in psychology, it is most commonly defined as the expectation that one has to justify their decisions and/or actions to others, typically to an authority (Lerner & Tetlock, 1999, Simonson & Nye, 1992). In addition, it usually entails that at least some direct or indirect consequences are contingent on the evaluation of these decisions by others (cf. Mulgan, 2000).

Accountability is integral to real-world decision-making, as "people do not think and act in a social vacuum" (Lerner & Tetlock, 1999, p.270) when they make decisions. Instead, often times our decisions are influenced and shaped by the anticipation that we, in some way, will be held accountable for them. Accountability is not only common in social settings, but it also plays an essential role in them. Because without accountability for our decisions and actions, we might act way more brashly than we should (cf. Hall et al., 2017). Therefore, accountability acts as a structural feature in our social environments to increase the thoroughness of our decisions and thus improve decision quality.

However, accountability is usually not understood as a unitarian concept but rather as contingent on several contextual factors. This understanding has been formalized in the *social contingency model of accountability* (Tetlock, 1985; Tetlock, 1992; Tetlock & Lerner, 1999). This theoretical framework, which, thus far, has been the most influential concept in accountability research, assumes that people's actions and decisions when they are being held accountable are driven mainly by two motives that compete with each other, namely *seeking audience approval* and *minimizing mental effort* (Tetlock & Lerner, 1999). Accordingly, people will act and decide in ways that they can easily defend against the scrutiny of others

while trying to use as little cognitive resources as possible. Based on these assumptions, the model proposes different coping strategies that people use to deal with being held accountable and that have been found in experimental research on accountability. The three most prominent of these coping strategies are *strategic attitude shifting* (decision-makers adjusting their public attitudes towards the views of the audience), *defensive bolstering* (self-justifying one's own decisions when confronted with a critical audience), and *preemptive self-criticism* (thinking about one's own decision thoroughly in order to anticipate potential criticism by the audience one is accountable to) (Tetlock & Lerner, 1999). These strategies affect the quality of judgment and decision-making in different ways. Strategic attitude shifting could have either a positive, a negative, or no effect on the quality of judgments and decisions. How strategic attitude shifting affects decisions depends on whether or not the preferred option of the audience one is accountable to is better, worse, or the same as the decision option that one would have chosen independently. This is because strategic attitude shifting causes the people that are held accountable to choose the option that the audience prefers. Defensive bolstering should generally have either no or a negative effect on decision and judgment quality, as people who engage in defensive bolstering make the same decisions they would have made without being held accountable. They only exert mental effort to defend their decision after they have already made it. However, by expanding cognitive effort to justify the decision one made retrospectively, the chance that one would realize that this decision was incorrect or suboptimal and therefore change it might be reduced. On the other hand, preemptive self-criticism should generally positively affect the quality of judgment and decision-making, as it should lead to accountable people making more thorough decisions by considering more different options and assessing these options more critically.

In their influential review of the accountability literature, Lerner and Tetlock (1999) identified four key moderators that affect which coping strategy is triggered in people when they are held accountable. The first moderator is knowledge about the views of the audience

one is accountable to, i.e., whether the views of the audience are known or unknown to the person that is being held accountable. When the audience's views are known, the social contingency model of accountability predicts that the person held accountable will aim to match these views with their decisions to gain the audience's social approval. This prediction is corroborated by a number of findings that show that people shift their attitudes towards that of an audience with known views with which they have to discuss their views (for example, Tetlock, 1983; Tetlock et al., 1989). In contrast, if one is accountable to an audience with unknown views, simply matching the audience's views is not possible. Therefore, it is believed that, in such a situation, accountability would lead to preemptive self-criticism (Lerner & Tetlock, 1999). Evidence for this has been provided, for example, by Tetlock (1983), who found that participants who were accountable to an unknown audience had a higher tolerance for evaluative inconsistency and recognized value-trade-offs better than participants that were accountable to a known audience when they evaluated controversial issues.

The second important moderator identified in the social contingency model of accountability is the timing of the accountability – whether a person learns that they are being held accountable before or after they made their decision. Post-decisional accountability has been shown to cause self-justification and defensive bolstering in individuals. Suppose one learns they are being held accountable after they have already made a decision. In that case, they will engage in effortful thinking to find post hoc justifications for their decision to defend it to others. This has been demonstrated, for example, by Conlon and Wolff (1980) as well as Fox and Staw (1979). Both of these studies found that post-decisional accountability increased escalating commitment. Pre-decisional accountability, on the other hand, should lead to preemptive self-criticism and, thus, have beneficial effects on decision quality. An example of this is the finding that pre-decisional accountability can reduce escalating commitment (Brockner et al., 1979; Simonson & Nye, 1992).

A third moderator proposed by the social contingency model of accountability is the legitimacy of accountability. While legitimate authorities should lead to positive and, therefore, potentially beneficial reactions to being held accountable (Tyler, 1997), illegitimate accountability demands should be perceived as intrusive and should cause negative reactions to these demands, warding off any positive effects accountability could, in theory, have (cf. Lerner & Tetlock, 1999).

Finally, the fourth moderator identified by Lerner and Tetlock (1999) is the accountability focus – whether one is predominantly held accountable for the outcomes (results) of their decisions or the process of their decisions (i.e., how the decisions were made). This fourth moderator has probably been the most influential in the accountability literature, and it will also be the main focus of this doctoral thesis. Therefore, I will discuss it in detail in the following section.

## **1.2 Process and outcome accountability**

According to the accountability literature, one of the most important and influential moderators for the effects of accountability on judgment and decision quality is the focus of accountability – that is, whether one is being held accountable for the *process* of their decision-making or the *outcome* of their decisions (for a review see Patil et al., 2014).

For people being held accountable for their decision-making process, it does not matter whether or not their decisions ultimately turn out to be correct. It only matters that they used the best or, at least, a justifiable procedure to come up with their decisions. One example of process accountability would be physicians on an oncology ward: a physician treating cancer would, in general, be held responsible for following the correct procedure (e.g., treating the patient through radiation or chemotherapy rather than treating them with globuli or prayers) instead of the health outcome per se, i.e., whether the patient lives or dies.

In contrast, outcome accountability describes a form of accountability that, as the name implies, is solely focused on the outcomes people achieve. Therefore, someone who is outcome accountable will be held accountable for the quality or the result of their decisions, regardless of the procedure they used to make this decision. In other words, in the case of outcome accountability, it does not matter whether or not someone used the ideal decision-making procedure or, at least, a rational one. It only matters whether that decision turned out to be good or not. An example of outcome accountability would be a CEO who is, for the most part, only held accountable for corporate profits. It only matters whether or not the company is profitable, not how the decisions were made that lead to the company being or not being profitable.

In general, empirical research on accountability has emphasized the benefits of process accountability on judgment and decision quality. For example, it has been shown that process accountability can improve interview validity (Brtek & Motowidlo, 2002), negotiation outcomes (De Dreu et al., 2000), and judgment accuracy (Siegel-Jacobs & Yates, 1996). Outcome accountability has not been associated with similar beneficial effects, with a few notable exceptions (de Langhe et al., 2011; Davis et al., 2007). If anything, outcome accountability has more often than not been characterized as detrimental to judgment and decision quality (e.g., Arkes et al., 1986; Siegel-Jacobs & Yates, 1996).

Accountability research has proposed two mechanisms to explain these differential effects of process and outcome accountability. The first one is stress: Siegel-Jacobs and Yates (1996) proposed that process and outcome accountability could affect stress levels differently. Outcome accountability has been claimed to increase stress to super-optimal levels (Siegel-Jacobs & Yates, 1996), which have been found to be detrimental to judgment and decision quality (Keinan et al., 1987; Yates, 1990). Process accountability, on the other hand, has been

associated with moderate increases in stress levels which, in turn, have been found to benefit the quality of judgment and decision-making (Janis & Mann, 1977; Yates, 1990).

The second mechanism proposed as an explanation for the differential effects of process vs. outcome accountability is epistemic motivation. Epistemic motivation is the desire to gain an accurate and detailed understanding of the world or, at least, the task or situation at hand. Epistemic motivation should, in turn, increase one's willingness to engage in effortful and systematic information processing. Whereas the stress explanation has mostly remained pure theoretical speculation, at least some studies reported evidence in favor of the predicted association of process accountability and epistemic motivation (De Dreu et al., 2006; Kruglanski, 1989). In contrast, outcome accountability has not been associated with increased epistemic motivation (De Dreu et al., 2000; De Dreu et al., 2006; de Langhe et al., 2011).

### **1.3 The problem with accountability research**

Research on accountability has made strong claims regarding the benefits of process accountability, both in comparison with no accountability and with outcome accountability, with the latter even being seen as a detriment to high-quality decision-making. Multiple influential reviews and textbooks (e.g., Bovens et al., 2014; Lerner & Tetlock, 1999; Schneider & Shanteau, 2003) have included and propagated these claims. However, upon close inspection, it becomes clear that these claims are based on comparatively little data. A lot of, even the most prominent, effects of process vs. outcome accountability (e.g., Simonson & Staw, 1992; Siegel-Jacobs & Yates, 1996) have been shown only once or twice under controlled conditions and, for a long time, independent replication studies have been entirely absent from this literature. This is a major caveat to the trustworthiness of these findings.

Philosophers of science, for the better part of a century, have argued that science without replicability is not science after all. An effect should only be assumed to be "real" if it can be repeated (Popper, 1959/2002) and if the procedures and circumstances that are



required to (re)produce the effect are known (Dunlap, 1926). These arguments have been substantiated by Ioannidis (2005), whose mathematical analyses have shown that any scientific field that neglects replications will end up in a situation where "most published research findings [in that field] are false" (Ioannidis, 2005, p.696). Based on this, it follows that "[...] direct replication by other laboratories is the best (and possibly the only) believable evidence for the reliability of an effect" (Simons, 2014, p.76). If such independent, direct replications do not exist, at least some level of skepticism regarding the robustness of the results of a study is warranted.

This skepticism towards the findings in the accountability literature is warranted further by the fact that psychology, as a scientific discipline, has been troubled in recent years by the so-called "replication crisis" (Pashler & Wagenmakers, 2012)<sup>1</sup>. There are a lot of different reasons for how and why this crisis in psychology came to be, ranging from systematic problems such as publication bias (Ioannidis, 2008; Pashler & Harris, 2012) that results in researchers leaving studies that did not find significant effects in their file-drawer instead of publishing them (Rosenthal, 1979), to researchers consciously or unconsciously using questionable research practices to achieve significant results (John et al., 2012; Simmons et al., 2011), and even scientific fraud (Stroebe et al., 2012). However, regardless of what the reasons for this crisis are in detail, or which of these reasons may be the most influential, one thing is obvious – and that is the fact that (lack of) replicability, indeed, is a significant problem in psychology. Multiple large-scale, multi-lab replication projects (e.g., Ebersole et al., 2016; Klein et al., 2014; Klein et al., 2018) have shown that replicability in psychological research, in general, is somewhat low. A recent review (Nosek et al., 2022) that

---

<sup>1</sup> This is not to say that replicability is only a problem in psychological research or that this problem particularly bad. It might very well be that this problem is just as bad or even worse in other disciplines. Instead, this statement aims to emphasize that replicability is a *known* problem in psychology. In contrast to this other disciplines instead of not having a replication crisis might simply be "a step behind" in that they have not identified and addressed the problem of low replicability in their field yet.

integrated the results of a large number of replication studies conducted between 2014 and 2021 has shown that only about 64% of replications reported statistically significant effects in the same direction as the effect in the original study. This means that less than 2/3 thirds of findings can be reliably replicated (Nosek et al., 2022). Furthermore, even in cases where effects can be replicated, they often turn out to be substantially smaller in the replication study as compared to the study where they were initially reported (Nosek et al., 2022).

In addition to the fact that most of the empirical research on accountability has not been replicated so far and, thus, is based on relatively little data, recent studies have called into question the generalizability of the superiority of process accountability compared with outcome accountability or no accountability. Most importantly, these recent studies have found no beneficial effects of process accountability. For example, Häusser et al. (2017) found no effect of process accountability on idea generation. Similarly, Hoffmann et al. (2017) did not find any effect of process accountability on judgment accuracy, and Chang et al. (2017) found that process accountability did not affect forecasting accuracy.

These findings that question the assumption of the general superiority of process accountability over outcome accountability and its general effectiveness at improving judgments and decision-making have been further corroborated by a new meta-analysis by Sharon et al. (2022). In this meta-analysis, the authors found that it is highly dependent on task characteristics, such as task complexity, which form of accountability (process or outcome accountability) is more conducive to good decision-making and general task performance.

However, the most concerning problem in terms of the robustness of accountability research is that recently several unsuccessful replication attempts were conducted, which were unable to replicate some of the beneficial effects of process accountability and the lack of such beneficial effects of outcome accountability that had been previously reported in the

literature. For instance, in the only direct replication study published yet, Schulz-Hardt et al. (2021) tried to replicate and extend the seminal work of Simonson and Staw (1992) but were unable to do so. The authors found no evidence that process (or outcome) accountability affected the amount of escalating commitment participants showed. Even when they integrated their findings with the original findings of Simonson and Staw (1992) with a Bayesian meta-analysis, they found substantial evidence for the null hypothesis that process accountability did not affect escalating commitment. Another unsuccessful replication of the beneficial effects of process accountability can be found in the dissertation of Sluga (2021). He was unable to replicate the very influential study by Siegel-Jacobs and Yates (1996). Finally, in a series of experiments conducted by Häusser and Schulz-Hardt (2017) as part of the DFG-project "Prozess- und Ergebnisverantwortlichkeit bei Gruppenentscheidungen ", the authors found that they were unable to replicate findings by De Dreu et al. (2000), Girandola and Gauthier (2001) and Scholten et al. (2007).

To summarize, accountability research, especially regarding the differential effects of process and outcome accountability, has been rather influential. It has changed our general understanding of human judgment and decision-making and shaped beliefs in how management systems should be designed in practice to facilitate sound decision-making. Yet, many of the claimed benefits of process accountability seem to be based on a comparatively small data basis, as many effects have been shown only once and lack any independent, direct (or conceptual) replications. In light of the general replication crisis in psychology, this alone should be cause for concern. However, even more concerning is a recent string of published (Schulz-Hardt et al., 2021) and unpublished replication attempts (Sluga, 2021; Häusser & Schulz-Hardt, 2017) that could not replicate the original findings. Considering this concerning set of circumstances surrounding the research on the effects of process and outcome accountability, the research presented in this doctoral thesis was devised and conducted.

The experiments that I will summarize in the following sections aim to further the scientific progress in the research field of accountability in two ways. The first manuscript focuses on advancing our knowledge about which effects from the accountability literature do replicate. To this end, three experiments were conducted to attempt to replicate and potentially expand another classic finding from the accountability literature: Vieider's (2009) findings that accountability can reduce loss aversion. While the first manuscript provides additional information regarding the extent of the replicability problem in accountability research, the second manuscript aims to gain insights into why findings on the effects of process vs. outcome accountability have failed to replicate. This was done by investigating some potential methodological problems in the literature by testing the effectiveness of the most common manipulation used to induce accountability in the laboratory.

## **2. Summary of Manuscript 1: "Does holding decision-makers accountable make them less loss averse? A three-study replication attempt."**

The initial aim of the research conducted for this manuscript was to extend a classic finding from the accountability literature. Namely, we tried to extend Vieider's (2009) finding that loss aversion can be reduced by holding people accountable. However, as it became apparent that we could not extend or even replicate Vieider's (2009) findings, the focus shifted towards a full-fledged, direct replication attempt.

Loss aversion is a very persistent cognitive bias that describes the tendency to weigh losses heavier than similarly sized gains (Tversky & Kahnemann, 1991) and can cause suboptimal decision-making, as it leads people to make irrational decisions in order to avoid losses (e.g., Kahneman et al., 1991; Köbberling & Wakker, 2005; Thaler, 1980). Vieider (2009) found that loss aversion can be reduced by holding people accountable for the decisions they make. Therefore, it would be quite beneficial if this finding was robust.

Specifically, this would mean that we have a tool at our disposal that allows us to reduce or even eliminate the effects of loss aversion and, thus, improve decision-making substantially.

In a series of three experiments, we closely followed the procedures Vieider (2009) used in his original study. Just like Vieider (2009), we told participants that they would be interviewed by a researcher after they completed working on the experimental task in order to manipulate accountability. They then worked on the same task as in Vieider (2009), the straight matching task, which was based on stimuli developed by Tversky and Kahnemann (1991). In this task, participants have to fill in hypothetical lotteries so that they would be indifferent to playing the lottery or maintaining the status quo (i.e., not playing the lottery). The lotteries are constructed in such a way that participants would win or lose the lottery with a 50% chance each, and the amount they could lose already filled in. Participants then only filled in the amount they would want to be able to win (gain) in order to be indifferent towards playing or not playing the lottery. We calculated their average loss aversion based on the gain values participants filled in.

While we generally followed Vieider's (2009) procedure very closely, we deviated in one major way from it in the first two experiments. The manipulation Vieider (2009) described in his study was somewhat ambiguous with regard to which form of accountability it was supposed to induce (although, at least in our opinion, it is much closer to a typical process accountability manipulation). In order to cleanly differentiate between the effects of process and outcome accountability on loss aversion and thus refine the specifics of the effect Vieider (2009) originally reported, we decided to manipulate accountability in a one-factorial design with three levels (no accountability vs. process accountability vs. outcome accountability) instead of Vieider's one-factorial two-level design (no accountability vs. accountability). The Bayesian analyses that we used provided substantial evidence for the null hypothesis. That is, they provided substantial evidence for the hypothesis that there are no

differences in loss aversion between process-accountable and non-accountable participants, neither in the first nor in the second experiment. Our analyses further provided evidence that there are also no differences in loss aversion between outcome-accountable and process-accountable participants or between outcome-accountable and non-accountable participants (though the evidence in the first experiment would have to be labeled as being only anecdotal).

As our first two experiments found no evidence at all of any effect of accountability (neither process nor outcome accountability) on loss aversion and therefore contradict the findings of Vieider (2009), we decided to conduct the third experiment as a full-fledged, direct replication. Thus, in our third experiment, we resorted to the one-factorial two-level design (accountability vs. no accountability) Vieider (2009) originally used in his study. In this third experiment, we also used a manipulation that resembled the one Vieider (2009) used as closely as possible. Furthermore, we also ensured that our sample's characteristics matched Vieider's (2009) in important aspects. Like Vieider (2009), in his original study, we collected a student sample with an academic background in business or economics and was fairly balanced in terms of gender composition. However, even with these changes, we could not replicate Vieider's (2009) findings in our third experiment. A Bayesian analysis again provided substantial evidence in favor of the null hypothesis (no differences in loss aversion between accountable and non-accountable participants). After all three of our experiments were unsuccessful in replicating the effect of (process) accountability on loss aversion that Vieider (2009) had found, we decided to integrate the available data in the form of a subsequent Bayesian meta-analysis. This meta-analysis, which combined the effect Vieider (2009) had found with the effects we found, again, provided substantial evidence in favor of the null hypothesis. From this, we can conclude that based on the currently available data for this research question, there is substantial evidence that (process) accountability does not affect loss aversion.

In conclusion, these results strongly indicate that the effect of (process) accountability on loss aversion that Vieider (2009) reported is far from being robust. Either there is no effect of (process) accountability on loss aversion at all and, thus, the original finding was a type I error, or the effect is far more volatile and context-dependent than previously thought. In either case, these findings considerably diminish the practical and theoretical implications that can be drawn from the originally reported effect. Additionally, these findings further question the general conclusion in the accountability literature that process accountability is superior to outcome accountability with regard to debiasing and improving judgment and decision-making. Combined with the other problematic findings described in the previous section (e.g., Schulz-Hardt et al., 2021; Sharon et al., 2022; Sluga, 2021), these results imply that the beneficial effects of process accountability have been overstated. Either the advantages of process accountability do not exist or, at the very least, they are not universal but instead highly situation- and context-specific. Therefore, it seems paramount that future research on accountability tries to identify such situational and contextual variables that moderate the effects of process and outcome accountability. Even more importantly, however, these findings further highlight the need for more scrutiny in the research of accountability, especially in the form of more direct and, ideally, large-scale replication attempts.

### **3. Summary of Manuscript 2: "What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms."**

Based on the findings from our first manuscript in conjunction with the other unsuccessful replication attempts that were recently reported (i.e., Häusser & Schulz-Hardt, 2017; Schulz-Hardt et al., 2021; Sluga, 2021), it has become increasingly clear that there is a serious problem of non-replicability in accountability research. We tried to understand what factors might play a role in causing these unsuccessful replications. In the manuscript

summarized in the following, we investigate the possibility that one of the most common manipulations used in lab experimental accountability research might be a substantial factor for the failed replications. This manipulation consists in announcing an interview to study participants in which they have to justify their decision-making process or their decision outcomes, depending on whether it is used to manipulate process or outcome accountability. The idea that this manipulation might be the reason behind the unsuccessful replication attempts is lent weight by the fact that all of the studies that have failed to replicate so far have used exactly this manipulation to induce accountability (De Dreu et al., 2000; Girandola & Gauthier, 2001; Scholten et al., 2007; Siegel-Jacobs & Yates, 1996; Simonson & Staw, 1992; Vieider, 2009). We identified two major problems with this interview manipulation. First, it appears to be rather weak and not very realistic, as people that are being held accountable in the laboratory via this interview manipulation are typically accountable for decisions that have little to no importance to them or anybody else. The decisions in the laboratory are either almost entirely inconsequential (e.g., multi-cue judgment tasks) or even purely hypothetical in nature (e.g., studies on escalating commitment or loss aversion). Besides this, while accountability in real-life settings often has severe consequences in the case of bad decisions (e.g., losing one's job or approbation), the consequences the interview manipulation entails for suboptimal or non-satisfactory decisions are practically neglectable. The worst consequence participants in the laboratory usually have to fear is the unpleasantness of explaining and justifying their suboptimal decisions to an expert who negatively reacts to this justification. Based on this, it would not be very surprising if the interview manipulation would indeed be too weak to induce differential effects of process and outcome accountability reliably.

Second, in many studies, the effectiveness of the interview manipulation is not appropriately controlled for. Many studies that use the interview-manipulation to induce accountability employ only minimal manipulation checks (e.g., Brtek & Motowidlo, 2002;



Chang et al., 2013; Siegel-Yacobs & Yates, 1996) or no manipulation checks at all (e.g., Hoffmann et al., 2017; Kim & Trotman, 2015; Lefebvre & Vieider, 2013; Vieider, 2009). Very often, it remains unclear whether or not the interview manipulation had any of the relevant psychological effects it is supposed to cause.

Therefore, we conducted two experiments in which we investigated the reliability of the interview manipulation as an effective manipulation of process and outcome accountability. In both experiments, we manipulated process and outcome accountability via the usual interview manipulation, and we compared both conditions with a control condition without accountability. In order to provide an in-depth check of the interview manipulations, we followed a three-step procedure in both experiments. In the first step, we test whether or not participants were aware of the manipulation and understood it correctly (instruction check). In the second step, we tested whether the manipulation induced feelings of process and outcome accountability, respectively. Finally, in the third step, we checked whether the interview manipulation activated the proposed psychological mediators for the effects of process and outcome accountability on judgment and decision quality, namely either differential effects of process vs. outcome accountability on psychological stress (Siegel-Jacobs & Yates, 1996) or an increase in epistemic motivation as a consequence of process accountability (Brtek & Motowidlo, 2002; De Dreu et al., 2006; de Langhe et al., 20110).

The first of the two experiments was conducted near the height of the global Covid-19 Pandemic, and since we aimed to achieve high test power, it was conducted as an online study. This is because it was impossible to collect large samples in the laboratory at that time. High test power was essential to this work in order to ensure that any failure to find substantial effects of the manipulation would be a consequence of a lack in the effectiveness of the manipulation rather than a type-II error. However, one problem that arose from the decision to conduct this first experiment online is that the results of it are not perfectly

comparable to those of other accountability studies since online studies are relatively uncommon in accountability research. In order to show that the results (or lack thereof) found in Experiment 1 of this manuscript were not only a result of it being conducted online, we replicated it with a nearly identical second experiment that was conducted as a classic laboratory study.

The results of both experiments were very similar. We found that, in both experiments, the interview manipulation was only partially understood by and salient to our participants. Specifically, we found that participants in the two accountability conditions correctly identified the contents of the interview announced to them above chance levels. However, a substantial number of participants were unable to do so (about 30% in the process accountability condition and almost 50% in the outcome accountability condition). This was especially surprising as participants had to answer this question (for the first time) very shortly after reading the instructions that included the experimental manipulation. Furthermore, participants in the process accountability condition believed it more likely that they would have to justify their decisions. In contrast, participants in the outcome accountability condition and participants in the no accountability condition assigned similar likelihoods to the possibility that they would have to justify themselves. Despite this, we found that the interview manipulation was, for the most part, successful in inducing feelings of process accountability in participants. However, it was less consistently successful in inducing felt outcome accountability, as it did so only in the first of the two experiments reported in this manuscript. Additionally, and perhaps most importantly, we found that the interview manipulation did not affect either stress or epistemic motivation. Lastly, supplementary analyses showed that process and outcome accountability had no effect on the judgment quality in the multi-cue estimation task that we used in the two experiments of this manuscript.

These findings show that the ability of the most common accountability manipulation to produce the psychological effects it is supposed to induce reliably is very questionable. It seems that the interview manipulation we investigated in this manuscript is either too weak to produce the expected effects consistently or that its effectiveness depends on contextual factors that have not yet been identified. Therefore, it stands to reason that the weakness or volatility of the interview manipulation might be the explanation (or at least one explanation) for the difficulties that other researchers and we had in replicating earlier findings on process and outcome accountability. A manipulation that only produces the anticipated effects under specific, *unknown* conditions would result in failed replications every time an experiment does not meet these specific conditions. In sum, our findings from this manuscript clearly demonstrate that accountability research needs to scrutinize its previous findings and assumptions further. In addition, these results also indicate that future research on accountability should look for new and better ways to induce process and outcome accountability.

#### **4. General Discussion**

The two manuscripts summarized above further the research on accountability in two important ways. For one, our findings clearly indicate that research on the effects of process and outcome accountability on decision-making has a serious problem with its replicability. Hence, the current consensus that process accountability is generally beneficial to judgment and decision-making is no longer warranted. Several recent studies (Häusser & Schulz-Hardt, 2017; Schulz-Hardt et al., 2021; Sluga, 2021) have already indicated that core findings of the literature on process and outcome accountability cannot be replicated. As shown in three experiments of the first manuscript summarized above, this list has to be extended. It no longer seems justified to claim that (process) accountability reduces loss aversion (as reported by Vieider, 2009). Furthermore, we did not find any beneficial effects of process

accountability on judgment accuracy in the two experiments of the second manuscript. This raises additional questions regarding the generalizability and replicability of previous findings reported by Siegel-Jacobs and Yates (1996) and de Langhe et al. (2011). Both these studies also used multi-cue judgment tasks and, in contrast to the two experiments in Manuscript 2, found differential effects of process and outcome accountability on judgment accuracy. Of course, the two experiments of Manuscript 2 cannot be considered direct replications of either Siegel-Jacobs and Yates (1996) or de Langhe et al. (2011). However, they still could be considered conceptual replications as one would expect to find the effects reported by these two studies to be present even if one uses a different multi-cue judgment task. The second significant contribution of the current set of experiments to the accountability literature is that the findings from the second manuscript provide the first substantial evidence for *why* so many of the recent accountability studies led to null findings. The findings from the second manuscript suggest that one major crux when trying to study accountability is the question of how to induce process and outcome accountability effectively and that the current way of doing that might be flawed. In the following, I will review these implications further, discuss the limitations of our research, and provide some ideas and directions for future research.

#### **4.1 Implications**

In all of our experiments, we found no indication that process accountability had any of the beneficial effects or that outcome accountability had any of the detrimental effects on judgment and decision quality that have been reported in the literature. The systematic problem regarding the replicability of earlier accountability research can have a number of reasons.

The first and most far-reaching possible reason that would cause the replications to fail is that the theoretical predictions regarding the effects of process and outcome accountability are simply wrong. In other words, different forms of accountability do not affect judgment

and decision-making differently, or maybe do not even affect them at all. All previous findings of such beneficial or detrimental effects of process and outcome accountability would then be type-I errors. However, since (replication) experiments always test more than just a theory alone (they always test a combination of the theory, the methods used, and a number of additional assumptions), assuming that all previous findings of beneficial effects of process accountability were type-I errors might be an overstatement and an overinterpretation of the evidence and, probably, not warranted at this point.

Another possible explanation for the lack of successful replications would be that the effects of process and outcome accountability are more context-dependent than previously thought, and the social contingency theory of accountability, therefore, is incomplete. This would mean that the failed replications are not indicative of previous findings being a type I error but rather show that the theory was incomplete and the procedures and assumptions to produce the effects reported in previous studies were insufficient. Two things speak for this explanation. First, as discussed earlier, the social contingency theory of accountability already identified additional moderators other than the accountability focus (process or outcome accountability), namely pre- vs. post-decisional accountability, legitimate vs. illegitimate accountability, and accountability to an audience with known vs. unknown views. While these other moderators are seldomly investigated anymore and generally are held constant in the research on process and outcome accountability (usually process vs. outcome accountability is investigated in the context of pre-decisional, legitimate accountability to an audience with unknown views), usually, there are no controls for this. Meaning studies researching process and outcome accountability generally are designed in a way that the accountability is pre-decisional and should be perceived by the study participants to be legitimate and towards an audience with unknown views. However, these studies practically never verify, for example, by using manipulation checks, whether participants feel that the audience they are accountable to is legitimate or whether they feel they know the audience's views. Accordingly, it would be

possible that interactions between the accountability focus and the other moderators affected the results of some of the reported studies, regardless of whether they were original findings or replication attempts. There is also another reason why it seems plausible that the problems replicating earlier findings stem from additional moderators that were not accounted for thus far. The social contingency theory of accountability does not claim that the list of four moderators I just named is exhaustive. Consequently, it is easily imaginable that more moderators exist that alter which effects accountability causes. This idea is supported by the findings of a recent meta-analysis by Sharon et al. (2022). In this meta-analysis, Sharon et al. (2022) found that the effects of process and outcome accountability were moderated by task complexity. Sharon et al. (2022) thus demonstrated that at least one more moderator exists that affects what consequences being held accountable has regarding judgment and decision quality.

Now, based on the findings in our second manuscript, there are at least two more explanations for the concerning number of unsuccessful replication attempts. As discussed previously, all of the studies that failed to replicate had one thing in common, they all used the aforementioned interview manipulation of accountability. This manipulation has two primary problems: One, it has relatively low ecological validity, and two, it seems rather weak as it entails little to no consequences for the study participants. In the two experiments of the second manuscript, we demonstrated that the interview manipulation of process and outcome accountability does not reliably activate the psychological mechanisms that should mediate the effects of process and outcome accountability, based on the previous accountability literature. Thus, a third possible explanation for the null findings of the replication attempts is that the interview manipulation used in both the original studies and the replication attempts is ineffective at producing the effects of process and outcome accountability. This would mean that the original findings obtained using the interview manipulation were false positives. However, this leaves the door open for the general theory on the effects of process and

outcome accountability to be correct. Since there are a decent number of studies that manipulate accountability in different ways, one could still assume that the effects of process and outcome accountability are real. However, only those that were found using a manipulation other than the interview manipulation. These other manipulations work by, for example, making rewards dependent on scores awarded to participants' decision-making process or decision outcomes (e.g., Chang et al., 2017; Tunguz & Carnevale, 2011) or by asking participants to write a statement explaining their decision-making process or decision outcomes (e.g., Chang et al., 2013; Dalla Via et al., 2019). Multiple studies that used these or similar manipulations of accountability also found beneficial effects of process accountability. For example, Libby et al. (2004) found that process accountability (manipulated by asking participants to write a statement in which they justified their procedures) can reduce the common measures bias (Libby et al., 2004), while Chang et al. (2013) found that it improved negotiation outcomes and Dalla Via et al., (2019) showed that process accountability increased decision quality. There is one piece of information, though, that questions this idea. Two of the original studies that failed to replicate did not *only* use the interview-manipulation. Both, Siegel-Jacobs and Yates (1996), as well as de Langhe et al. (2011), manipulated outcome accountability by basing the payoffs participants could receive on the accuracy of their judgments. As both those studies still did not replicate (cf. Häusser & Schulz-Hardt, 2017; Sluga, 2021), it can be assumed that the problem of replicability in accountability research extends at least somewhat beyond studies using the interview manipulation.

A final possibility regarding the interpretation of our findings is that the effectiveness of the interview manipulation depends on specific context factors instead of being generally ineffective. It is possible that whether or not the interview manipulation is effective at inducing process and outcome accountability depends on a number of moderator variables that have not been identified yet. In theory, there is an almost infinite number of such possible context factors, but plausible candidates would be differences in the subject population,

cultural differences, subtle differences in the laboratory setting, or simply changes over time. It might be that while the interview manipulation is effective in the Netherlands, it is not in Germany because of the cultural differences between the two countries. Another possibility is that the manipulation was effective in the past but is ineffective today. It is impossible to rule out such hidden moderators fully. However, two things should be noted here. First, while such hidden moderators have been leveraged as an explanation for unsuccessful replication, recently, it has become increasingly clear that hidden moderators are not very common. The large-scale, collaborative replication project "Many Labs 2" recently showed that impactful hidden moderators that affect whether an effect is found or not are very rare (Klein et al., 2018). Second, even if such hidden moderators exist, the interview manipulation should be avoided, and conclusions drawn from studies using it should be scrutinized.

Given that it is unclear what these hidden moderators are, it is impossible for researchers trying to use the interview manipulation to determine ahead of time whether they achieved the correct set of contextual factors to make it effective. This would result in very inefficient research (and possibly also very large "file drawers"). While it is always the case that researchers only know whether their manipulation was effective after they ran a study, the likelihood that a manipulation that has been used previously turns out to be ineffective is relatively low in cases where no hidden moderators exist. If hidden moderators for the effectiveness of the manipulation do exist, however, the likelihood that the effectiveness of the manipulation varies from one study to another rises drastically. As long as the moderators are hidden, researchers do not know what factors they actually need to control for in order to achieve an effective manipulation. Furthermore, if the effects of the manipulation are so fragile that subtle situational differences affect them, it seems rather unlikely that effects identified with this research paradigm would generalize to real-world settings and are applicable in practice.



Regardless of which of the four options described above represents the state of reality the best, some things are clear. For one, previous findings regarding the effects of process and outcome accountability and conclusions drawn from these findings should be regarded with great caution, as one should be wary of the possibility that they might turn out to be wrong. This is especially true for studies that used the interview manipulation of accountability, and one should not rely on these findings to base practical recommendations on. Given how common the interview manipulation is in studies researching process accountability, this alone is enough to conclude that the general consensus in the accountability literature that process accountability has predominantly beneficial effects is not substantiated anymore. Furthermore, this also means that, in any case, future research should abandon using the interview manipulation altogether. Instead, researchers should use either one of the other accountability manipulations mentioned above, given that these turn out to be more reliable than the interview manipulation, or look for new research paradigms to induce accountability.

## **4.2 Limitations and Future Research**

In the work I summarized above, we were able to confirm that there is a clear and consistent problem concerning replicability in accountability research and that one possible reason for this problem lies in the use of a research paradigm that is unable to induce accountability and its mediating psychological mechanisms reliably. However, there are some limitations to these conclusions and some avenues for future research that I want to discuss in the following.

First, while the reported findings do, to some extent, further question the validity of the general claim that process accountability is superior to outcome accountability and non-accountability, they mainly demonstrate that the interview manipulation is problematic. Like most past replication attempts, this work again focused on experiments in which accountability was manipulated by announcing an interview, and like these previous replication attempts, we found no effects of the manipulation. From this follows, however,

that the conclusions that can be drawn from these findings are more or less limited to findings that were based on the interview manipulation. Our findings are not informative regarding the effects that were found using other manipulations of accountability. It is possible that these manipulations turn out to be unreliable as well and that the effects found with them are, therefore, not robust. However, it is equally likely that they are indeed effective, and thus, the results found using these manipulations are reliable. Therefore, it would be important to try to replicate findings from the accountability literature that used other manipulations than the interview manipulation. Such replication attempts would provide important additional information regarding the cause of the difficulties researchers had so far in replicating the effects of process and outcome accountability. On top of this, such replication attempts might also help by identifying more robust accountability manipulations that could be used moving forward.

Depending on whether or not such replication attempts of studies using different accountability manipulations can identify a robust accountability manipulation, another important direction for accountability research could be to develop new paradigms to manipulate accountability. Such a new paradigm should aim to be more lifelike by heightening the stakes involved and simultaneously increasing the consequences for subpar decisions and/or decision-making.

Another limitation of the work presented in this dissertation is that we cannot rule out that contextual factors, i.e., moderator variables that we are unaware of, caused some or all of the unsuccessful replication attempts. Neither previous replication studies nor the work presented above controlled for moderator variables, neither for the already known ones (accountability timing, knowledge about the audience's views, and legitimacy of accountability) nor for ones that are less established or new. Future research should therefore be more mindful of the known moderators of accountability effects and try to control that these indeed have the correct characteristics. Meaning in the future, researchers should

confirm not only whether or not process and outcome accountability were correctly induced but also whether the accountability was correctly perceived as being pre-decisional and towards a legitimate authority with unknown views. In addition, future research should also look to identify additional moderators that could affect the effects of process and outcome accountability, as there is, as discussed above, some indication that such additional moderators do exist (cf. Sharon et al., 2022).

Finally, one last limitation is the use of student samples. It is common practice in psychological research in general but also in accountability research specifically (cf. Aleksovskaja et al., 2019) to conduct studies with students as experimental subjects. However, this use of student samples has certain problems. University students differ in multiple ways from the general public, which can potentially reduce the generalizability of effects found in student samples. One way in which students differ from the general public is in terms of demographic characteristics, as they are younger than the average population of most countries, are less financially stable, and for the most part, are unmarried (Ashraf & Merunka, 2016). These demographic differences are not, per se, problematic for the ecological validity of psychological research, as these differences alone do not necessarily affect behavior. Such demographic differences are only impactful for the conclusions drawn from a specific demographic when the differences to other demographic groups are accompanied by psychological differences. Researchers have identified a number of such accompanying psychological differences between students and the general public. Sears (1986) has argued that students compared to the general public, have, among other things, weaker self-definition, less crystallized attitudes (which results in less consistent attitudes and causes students to be more susceptible to external influence), and better cognitive skills. These psychological differences, while in general somewhat peculiar for psychological research, are especially problematic for research, such as research on accountability that is, at least in the broadest sense, concerned with the effects social influence has on individuals. If the subjects

we study are more prone to be influenced than the general public, our research will often lead to overestimations of effects and might cause us to give practical recommendations that do not stand on solid empirical ground.

Regarding accountability research, this might turn out to be critical, as it is possible that the conclusions drawn from student samples do not generalize to the general population. It could very well be that while being held accountable affects the behavior of students, as they are comparatively easy to influence, it does not affect the behavior of the average adult, who has more crystallized attitudes and is generally harder to influence. Given that findings from accountability research are used quite often to give practical recommendations, for example, on how to design management systems in public and private organizations in order to improve decision-making in these organizations, this is rather concerning.

Therefore, one more, important avenue for future research should be to focus more on accountability in "the wild", meaning doing field studies on accountability. This would not only solve the problem of generalizability, but it would also address some of the other methodological issues I have already discussed, namely the difficulties laboratory studies face when they try to manipulate accountability. Field studies do not face these problems. In the field, accountability is "real", meaning that participants in field studies face real consequences for what decisions they made or for how they made their decisions. On top of this, the decisions made in the field involve, or at least can involve, much higher stakes, making the decisions more impactful all-around. Field studies on accountability thus help solve the problem of accountability manipulations in the laboratory being too weak. In addition, field studies investigate accountability in the same population they want to make statements about and give recommendations to. This drastically improves the ecological validity of findings and makes them directly applicable to practice.

## 5. Conclusion

Research on accountability has made bold claims on the superiority of holding people process-accountable compared to holding people accountable for outcomes or not accountable at all. These claims have found their way into influential textbooks, and reviews (e.g., Bovens et al., 2014; Lerner & Tetlock, 1999; Schneider & Shanteau, 2003) and are potentially used by practitioners in the organizational context to design management systems in order to optimize decision-making quality. However, just like many other fields in psychology, accountability research has, until recently, neglected replications in favor of constantly looking for new effects. This has resulted in a recent rise in questions regarding the robustness of the effects of process and outcome accountability. The research presented in this doctoral thesis provides further evidence that previous findings regarding the superiority of process accountability over outcome accountability are less robust than originally thought. In addition, it provides a possible explanation for why a number of earlier findings seem not to be replicable, as a thorough examination of one of the most common experimental manipulations found it to be anything but reliable. This work lays the foundation for future research to build on to make research on the effects of accountability more robust and, thus, trustworthy.

## 6. References

- Aleksovska, M., Schillemans, T., & Grimmelikhuijsen, S. (2019). Lessons from five decades of experimental and behavioral research on accountability: A systematic literature review. *Journal of Behavioral Public Administration*, 2(2).  
<https://doi.org/10.30636/jbpa.22.66>
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes*, 37(1), 93–110. [https://doi.org/10.1016/0749-5978\(86\)90046-4](https://doi.org/10.1016/0749-5978(86)90046-4)
- Ashraf, R., & Merunka, D. (2016). The use and misuse of student samples: An empirical investigation of European marketing research. *Journal of Consumer Behaviour*, 16(4), 295–308. <https://doi.org/10.1002/cb.1590>
- Bovens, M. A. P., Goodin, R., & Schillemans, T. (2014). *The Oxford Handbook of Public Accountability*. Oxford University Press.
- Brockner, J., Shaw, M. C., & Rubin, J. Z. (1979). Factors affecting withdrawal from an escalating conflict: Quitting before it's too late. *Journal of Experimental Social Psychology*, 15(5), 492–503. [https://doi.org/10.1016/0022-1031\(79\)90011-8](https://doi.org/10.1016/0022-1031(79)90011-8)
- Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *The Journal of Applied Psychology*, 87(1), 185–191.  
<https://doi.org/10.1037/0021-9010.87.1.185>
- Chang, Atanasov, Patil, Mellers, & Tetlock. (2017). Accountability and adaptive performance under uncertainty: A long-term view. *Judgment and Decision Making*, 12(6), 610–626.
- Chang, L. J., Cheng, M. M., & Trotman, K. T. (2013). The effect of outcome and process accountability on customer–supplier negotiations. *Accounting, Organizations and Society*, 38(2), 93–107. <https://doi.org/10.1016/j.aos.2012.12.002>
- Conlon, E. J., & Wolf, G. (1980). The moderating effects of strategy, visibility, and involvement on allocation behavior: An extension of Staw's escalation paradigm. *Organizational Behavior and Human Performance*, 26(2), 172–192.  
[https://doi.org/10.1016/0030-5073\(80\)90053-7](https://doi.org/10.1016/0030-5073(80)90053-7)

- Dalla Via, N., Perego, P., & van Rinsum, M. (2019). How accountability type influences information search processes and decision quality. *Accounting, Organizations and Society*, 75, 79–91. <https://doi.org/10.1016/j.aos.2018.10.001>
- Davis, W. D., Mero, N., & Goodman, J. A. (2007). The interactive effects of goal orientation and accountability on task performance. *Human Performance*, 20(1), 1–21. <https://doi.org/10.1080/08959280709336926>
- De Dreu, C. K., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: a motivated information-processing approach to integrative negotiation. *Journal of Personality and Social Psychology*, 79(6), 975–987. <https://doi.org/10.1037/0022-3514.79.6.975>
- De Dreu, C. K. W., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of Personality and Social Psychology*, 90(6), 927–943. <https://doi.org/10.1037/0022-3514.90.6.927>
- Dunlap, K. (1926). The experimental methods of psychology. In: *Psychologies of 1925* ed. C. Murchison, pp. 331–53. Clark University Press.
- Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., [...] Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, 67, 68–82. <https://doi.org/10.1016/j.jesp.2015.10.012>
- Fox, F. V., & Staw, B. M. (1979). The trapped administrator: Effects of job insecurity and policy resistance upon commitment to a course of action. *Administrative Science Quarterly*, 24(3), 449–471. <https://doi.org/10.2307/2989922>
- Girandola, F., & Gauthier, E. (2001). Organizational decision making: Effects of accountability on escalation of commitment. *European Review of Applied Psychology = Revue Europeenne de Psychologie Appliquee*, 51(1), 111–119.
- Hall, A. T., Frink, D. D., & Buckley, M. R. (2017). An accountability account: A review and synthesis of the theoretical and empirical research on felt accountability. *Journal of Organizational Behavior*, 38(2), 204–224. <https://doi.org/10.1002/job.2052>

- Häusser, J. A., Frisch, J. U., Wanzel, S., & Schulz-Hardt, S. (2017). Effects of process and outcome accountability on idea generation. *Experimental Psychology*, *64*(4), 262–272. <https://doi.org/10.1027/1618-3169/a000368>
- Häusser, J. A., & Schulz-Hardt, S. (2017). *Ergebnisbericht zum DFG-Projekt „Prozess- und Ergebnisverantwortlichkeit bei Gruppenentscheidungen“*. Universität Gießen & Universität Göttingen.
- Hoffmann, J. A., Gaissmaier, W., & von Helversen, B. (2017). Justifying the judgment process affects neither judgment accuracy, nor strategy use. *Judgment and Decision Making*, *12*(6), 627–641.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. <https://doi.org/10.1371/journal.pmed.1004085>
- Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5), 640–648. <https://doi.org/10.1097/EDE.0b013e31818131e7>
- Janis, I. L., & Mann, L. (1977). *Decision Making: A Psychological Analysis of Conflict, Choice, and Commitment* (Bd. 488). Free Press.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*(5), 524–532. <https://doi.org/10.1177/0956797611430953>
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, *5*(1), 193–206. <https://doi.org/10.1257/jep.5.1.193>
- Keinan, G., Friedland, N., & Ben-Porath, Y. (1987). Decision making under stress: Scanning of alternatives under physical threat. *Acta Psychologica*, *64*(3), 219–228. [https://doi.org/10.1016/0001-6918\(87\)90008-4](https://doi.org/10.1016/0001-6918(87)90008-4)
- Kim, S., & Trotman, K. T. (2015). The comparative effect of process and outcome accountability in enhancing professional scepticism. *Accounting and Finance*, *55*(4), 1015–1040. <https://doi.org/10.1111/acfi.12084>
- Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., Bocian, K., Brandt, M. J., Brooks, B., Brumbaugh, C. C., Cemalcilar, Z., Chandler, J., Cheong,



- W., Davis, W. E., Devos, T., Eisner, M., Frankowska, N., Furrow, D., Galliani, E. M., [...] Nosek, B. A. (2014). Investigating variation in replicability. *Social Psychology*, 45(3), 142–152. <https://doi.org/10.1027/1864-9335/a000178>
- Klein, R. A., Vianello, M., Hasselman, F., Adams, B. G., Adams, R. B., Alper, S., Aveyard, M., Axt, J. R., Babalola, M. T., Bahník, Š., Batra, R., Berkics, M., Bernstein, M. J., Berry, D. R., Bialobrzeska, O., Binan, E. D., Bocian, K., Brandt, M. J., Busching, R., [...] Nosek, B. A. (2018). Many Labs 2: Investigating variation in replicability across samples and settings. *Advances in Methods and Practices in Psychological Science*, 1(4), 443–490. <https://doi.org/10.1177/2515245918810225>
- Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of Economic Theory*, 122(1), 119–131. <https://doi.org/10.1016/j.jet.2004.03.009>
- Kruglanski. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*. <https://doi.org/10.1037/0033-2909.106.3.395>
- Langhe, B. de, van Osselaer, S. M. J., & Wierenga, B. (2011). The effects of process and outcome accountability on judgment process and performance. *Organizational Behavior and Human Decision Processes*, 115(2), 238–252. <https://doi.org/10.1016/j.obhdp.2011.02.003>
- Lefebvre, M., & Vieider, F. M. (2013). Reining in excessive risk-taking by executives: the effect of accountability. *Theory and Decision*, 75(4), 497–517. <https://doi.org/10.1007/s11238-012-9335-2>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Libby, T., Salterio, S. E., & Webb, A. (2004). The balanced scorecard: The effects of assurance and process accountability on managerial judgment. *The Accounting Review*, 79(4), 1075–1094. <https://doi.org/10.2308/accr.2004.79.4.1075>
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(6), 537–542. <https://doi.org/10.1177/1745691612460688>

- Mulgan, R. (2000). "Accountability ": An ever-expanding concept? *Public Administration*, 78(3), 555–573. <https://doi.org/10.1111/1467-9299.00218>
- Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Kline Struhl, M., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, 73, 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(6), 531–536. <https://doi.org/10.1177/1745691612463401>
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A Crisis of Confidence? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(6), 528–530. <https://doi.org/10.1177/174569161246525>
- Patil, S. V., Vieider, F., & Tetlock, P. E. (2014). Process versus outcome accountability. *The Oxford Handbook of Public Accountability*, 69–89. <https://doi.org/10.1093/oxfordhb/9780199641253.013.0002>
- Popper, K. R. (1959/2002). *The Logic of Scientific Discovery*. Psychology Press.
- Schneider, S. L., & Shanteau, J. (Hrsg.). (2003). Emerging perspectives on judgment and decision research. *Cambridge Series on Judgment and Decision Making.*, 713. <https://doi.org/10.1017/CBO9780511609978>
- Scholten, L., van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. W. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, 43(4), 539–552. <https://doi.org/10.1016/j.jesp.2006.05.010>
- Schulz-Hardt, S., Rollwage, J., Wanzel, S. K., Frisch, J. U., & Häusser, J. A. (2021). Effects of process and outcome accountability on escalating commitment: A two-study replication. *Journal of Experimental Psychology: Applied*, 27(1), 112–124. <https://doi.org/10.1037/xap0000321>

- Sears, D. O. (1986). College sophomores in the laboratory: Influences of a narrow data base on social psychology's view of human nature. *Journal of Personality and Social Psychology*, *51*(3), 515–530. <https://doi.org/10.1037/0022-3514.51.3.515>
- Sharon, I., Drach-Zahavy, A., & Srulovici, E. (2022). The effect of outcome vs. process accountability-focus on performance: A Meta-Analysis. *Frontiers in Psychology*, *13*. <https://doi.org/10.3389/fpsyg.2022.795117>
- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, *65*(1), 1–17. <https://doi.org/10.1006/obhd.1996.0001>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *9*(1), 76–80. <https://doi.org/10.1177/1745691613514755>
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P-Curve and Effect Size: Correcting for Publication Bias Using Only Significant Results*. <https://doi.org/10.2139/ssrn.2377290>
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, *51*(3), 416–446. [https://doi.org/10.1016/0749-5978\(92\)90020-8](https://doi.org/10.1016/0749-5978(92)90020-8)
- 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>
- Sluga, A. (2021). *Hour of Judgment: On Judgment, Decision Making, and Problem Solving under Accountability* (Nr. EPS-2021-520-F&A). <https://repub.eur.nl/pub/136967/hffinaluploaddissertation.pdf>
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science: A Journal of the*

- Association for Psychological Science*, 7(6), 670–688.  
<https://doi.org/10.1177/1745691612460687>
- Tetlock, P. E. (1983). Accountability and complexity of thought. *Journal of Personality and Social Psychology*, 45(1), 74–83. <https://doi.org/10.1037/0022-3514.45.1.74>
- Tetlock, P. E. (1985). Accountability: The neglected social context of judgment and choice. *Research in Organizational Behavior*, 7(1), 297–332.
- Tetlock, P. E. (1992). The impact of accountability on judgment and choice: Toward a social contingency model. In M. P. Zanna (Hrsg.), *Advances in Experimental Social Psychology* (Bd. 25, S. 331–376). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60287-7](https://doi.org/10.1016/S0065-2601(08)60287-7)
- Tetlock, P. E., & Lerner, J. S. (1999). The social contingency model: Identifying empirical and normative boundary conditions on the error-and-bias portrait of human nature. In S. Chaiken (Hrsg.), *Dual-process Theories in Social Psychology*, (pp (Bd. 657, S. 571–585). The Guilford Press, xiii.
- Tetlock, P. E., Skitka, L., & Boettger, R. (1989). Social and cognitive strategies for coping with accountability: conformity, complexity, and bolstering. *Journal of Personality and Social Psychology*, 57(4), 632–640. <https://doi.org/10.1037/0022-3514.57.4.632>
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39–60. [https://doi.org/10.1016/0167-2681\(80\)90051-7](https://doi.org/10.1016/0167-2681(80)90051-7)
- Tunguz, S., & Carnevale, P. J. (2011). Social context effects in the display of emotion: Accountability in a simulated organization. *Journal of Applied Social Psychology*, 41(6), 1371–1398. <https://doi.org/10.1111/j.1559-1816.2011.00768.x>
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061.  
<https://doi.org/10.2307/2937956>
- Vieider, F. M. (2009). The effect of accountability on loss aversion. *Acta Psychologica*, 132(1), 96–101. <https://doi.org/10.1016/j.actpsy.2009.05.006>
- Walker, R. M., James, O., & Brewer, G. A. (2017). Replication, experiments and knowledge in public management research. *Public Management Review*, 19(9), 1221–1234.  
<https://doi.org/10.1080/14719037.2017.1282003>

Yates, J. F. (1990). *Judgment and Decision Making* (Bd. 430). Prentice-Hall, Inc.

## Appendix A

### Manuscript 1

Does holding decision-makers accountable make them less loss averse? A three-study replication attempt.

**Does holding decision-makers accountable make them less loss  
averse? A three-study replication attempt.**

**Johannes Rollwage**

University of Goettingen

**Jan Alexander Häusser**

Justus-Liebig-University of Giessen

**Stefan Schulz-Hardt**

University of Goettingen

### Author Note

Johannes Rollwage & Stefan Schulz-Hardt, Institute of Psychology, Economic and Social Psychology Unit, University of Goettingen, Germany; Jan Häusser, Institute of Psychology, Justus-Liebig-University of Giessen, Germany.

The authors have no conflict of interest to disclose and have complied with APA ethical standards in human subjects research. Data has not been previously distributed. The full set of materials and data can be found in the OSF repository (<https://osf.io/f7wj9/>). The research reported in this manuscript was funded by research grants from the German Science Foundation (Deutsche Forschungsgemeinschaft, DFG) to the second and the third author, project nos. HA 6455/1-1 and SCHU 1279/14-1. The authors would like to thank Maximilian Blomberg, Arzu Çetin, Carlotta Dove, Moritz Ernst, Johanna Frisch, Maren Michaelis, Sonja Pilz, Tjark Rode, Jana Schildknecht, Hanna Scholz, Stella Wanzel, and Anja Wegner for their assistance in collecting the data of the three experiments reported here.

Correspondence concerning this manuscript should be addressed to Johannes Rollwage, Georg-August-University, Institute of Psychology, Economic and Social Psychology Unit, Gosslerstrasse 14, 37073 Goettingen, Germany. Email may be sent to [rollwage@psych.uni-goettingen.de](mailto:rollwage@psych.uni-goettingen.de).



## Abstract

People tend to value losses higher than equivalent gains. This phenomenon, aptly named *loss aversion*, can have detrimental effects on decision-making by, for example, making people avoid worthwhile options if there is even a small risk of a large loss. Vieider [(2009). The effect of accountability on loss aversion. *Acta Psychologica*, 132(1), 96–101.] reported that this irrational tendency to avoid losses can be reduced by holding people accountable. The present study initially aimed to clarify whether this effect is driven by either process or outcome accountability. However, as we did not find any effects of either process or outcome accountability on loss aversion in two experiments, our focus shifted towards a direct replication attempt. Our third experiment used materials, procedures, and manipulations that closely matched those of the original study, and almost doubled its sample size. Despite these efforts, we still did not replicate Vieider's findings, with a Bayesian test providing substantial evidence in favor of the null hypothesis of no difference between experimental conditions. In sum, our results question the robustness of the effect of accountability on loss aversion. In this, they join other recent studies that fail to replicate earlier findings regarding the effects of accountability on decision-making.

*Keywords:* loss aversion, process accountability, outcome accountability, replication

## **Does holding decision-makers accountable make them less loss averse? A three-study replication attempt.**

*Loss aversion* describes the tendency of people to weigh losses heavier than gains of the same magnitude (Tversky & Kahnemann, 1991) and is assumed to be an example of an even more general cognitive bias pervading the human condition, referred to as the *negativity bias* (Baumeister et al., 2001; Rozin & Royzman, 2001). Generally, loss aversion is believed to be a fundamental part of human nature rooted in our evolutionary origins, as studies have shown loss aversion in different cultures (Maddux et al., 2010), in children (Harbaugh et al., 2001), and even in non-human primates (Brosnan et al., 2007; Kanngiesser et al., 2011; Lakshminaryanan et al., 2008). While it is likely that loss aversion had some evolutionary fitness benefit in the past, it seems that this choice tendency has overgeneralized in humans today and reaches far beyond the domains of survival (Camerer, 2005). Examples of suboptimal or, at least, irrational decision-making caused by loss aversion include a diverse set of psychological phenomena like the endowment effect (Thaler, 1980; Tversky & Kahneman, 1991), the status quo bias (Kahneman et al., 1991, Samuelson & Zeckhauser, 1988;), the equity premium puzzle (Benartzi & Thaler, 1995; Thaler et al., 1997), the disposition effect (Odean, 1998; Shefrin & Statman, 1985; Weber & Camerer, 1998), asymmetries in price elasticity (Putler, 1992), and risk aversion (Köbberling & Wakker, 2005). In general, loss aversion can lead to irrational and detrimental decisions by making people either too averse to risks, e.g., by overvaluing moderate but certain earnings compared to very high but slightly uncertain earnings, or too eager to take risks, e.g., when they face a certain but small loss, they might take a risky gamble for the off-chance of avoiding a loss altogether. Therefore, it would be beneficial to identify ways in which loss aversion can be reduced and, thus, debias decision-making. Some studies have already provided evidence in this vein. For example, List (2004) found that loss aversion is reduced for people with more experience in market transactions. While market experience seems to be a successful tool to

reduce loss aversion, it is not necessarily easy to obtain, so identifying more applicable and ready-to-use interventions to reduce loss aversion is crucial.

One potential intervention to debias decision-making in general that has been prominent in the last few decades is *accountability*. Usually, decision-makers in organizations are held accountable for their decisions and, thus, have to justify themselves and/or their behavior to others. This is done to ensure that their decisions are made with the required thoroughness and usually entails that decision-makers who cannot offer a satisfactory justification suffer some negative consequences. This, in turn, should prompt decision-makers to be more diligent and thorough with the decisions they make and, thus, reduce decision biases (Lerner & Tetlock, 1999; Simonson & Nye, 1992).

In line with this idea, Vieider (2009) reported evidence that accountability reduces loss aversion. In a laboratory experiment, he manipulated the perceived accountability by informing half of the participants that they would be interviewed about their choices by an experimenter. The announcement of such interviews is the most common method for experimentally manipulating accountability in lab experiments (e.g., de Langhe et al., 2011; Girandola & Gauthier, 2001; Simmons & Staw, 1992). The other half of the participants received no such announcement. In contrast, they were informed that their choices could not be traced back to them and would be kept confidential. The decision task Vieider's subjects worked on was adapted from stimuli developed by Tversky and Kahnemann (1992). In this task, participants had to fill in a value into a two-outcome prospect (*simple prospect*) in such a way that they would be indifferent between the prospect and the status quo. The status quo represented a monetary outcome of 0, while the prospect made the participant lose a set amount of money with a probability of 0.5 or win the amount of money the participant had filled in with a probability of 0.5. Higher proposed gains (to reach a point of indifference)

reflect stronger loss aversion. Vieider found that accountable participants showed significantly less loss aversion than non-accountable participants.

To the best of our knowledge, no replications of this effect have been published so far. There is one study by Pahlke, Strasser, and Vieider (2012) where somewhat similar results have been reported, but with some important caveats. In this study, participants had to choose 40 times between a certain monetary outcome and a binary prospect, i.e., a gamble with two outcomes, both of which occur with a 50% chance. Pahlke and colleagues constructed these binary prospects as either gain (always have a positive expected value), loss (always have a negative expected value), or mixed (always have an expected value of 0) prospects. The amount of loss aversion was then measured by calculating the relative frequency of safe choices made for the mixed prospects. In order to manipulate accountability, Pahlke et al. randomly assigned participants to either a responsibility condition or an accountability condition, and either to the role of decision-maker or the role of recipient. The decision-makers made the 40 choices between the sure amount of money and the binary prospects for both themselves and the recipient, while the recipient only saw the choices the decision-maker made and indicated whether they were satisfied with the choice or not. In the responsibility condition, there was no direct interaction between the decision-maker and the recipient, as the identity of the decision-maker was not disclosed to the recipient. In contrast to this, in the accountability condition, decision-makers were informed that, after the end of the experimental session, one of them would be randomly extracted and would be interrogated by the recipient about the reasons behind their choices. Pahlke et al. (2012) found a significant interaction between accountability and mixed prospects, indicating that participants in the accountability condition choose the safe outcome for the mixed prospects more seldomly than participants in the responsibility condition. However, they did not find any significant interactions between accountability and gain or loss prospect, even though one would expect significant loss aversion (that should be reduced by being held accountable), at least for loss

prospects. These findings, while similar to those in Vieider's (2009) original study, therefore only partially corroborate them. Accordingly, only the original Vieider (2009) study provides clear evidence for the effect of accountability on loss aversion, which is why we will predominantly focus on it in our own experiments.

Despite all their other similarities and differences, there is one important caveat that both these studies (Pahlke et al., 2012; Vieider, 2009) have in common, and that is that they both neglect a crucial and common differentiation in accountability research. Namely, they both omit the differentiation between process accountability and outcome accountability, which are two different types of accountability assumed to exert different effects on judgments and decisions.

### **Process versus outcome accountability**

Differentiating accountability into process and outcome accountability has been an influential step in accountability research (for an overview, see Patil et al., 2014). Outcome accountability is marked by a focus only on the outcomes that decision-makers produce, i.e., when they are evaluated, they have to justify what outcomes they achieved, not how they got there. This is the kind of accountability usually found in economic organizations, e.g., the CEO of a company is held accountable for corporate profits.

In contrast to this, process accountability is a form of accountability where the opposite is the case, i.e., it only matters what kind of process a decision-maker used to reach a judgment or decision, but not what the outcome of that judgment or decision ultimately was. This kind of accountability is primarily found in public offices and the health care sector, e.g., the attending physician of a gravely ill patient might be held accountable if they administer a non-proven treatment even if the health outcome is positive, while they will not be held accountable if they administer a proven treatment even if the health outcomes are adverse. The main idea behind holding people process accountable is that outcomes cannot always be

guaranteed even if the most efficient strategy was used and, thus, decision-makers should only be held accountable for things that they can control.

The differentiation of these two types of accountability is important because, so far, most empirical (experimental) research on process versus outcome accountability indicates that these two types of accountability have very different effects on judgments and decisions: whereas, according to the literature, process accountability increases the quality of judgments and decisions, outcome accountability often has been reported to have detrimental effects. For example, Brtek and Motowidlo's (2002) found that process accountable participants made more valid predictions regarding the future job performance of 60 managers (based on simulated job interviews) than non-accountable participants did. In contrast to this, outcome accountable participants' judgments were less valid with regard to job performance than those of non-accountable participants. There are several studies that demonstrated similar differential effects of process and outcome accountability, such as differences in judgment quality (Siegel-Jacobs & Yates, 1996), or differences in the revision of the fixed-pie bias in negotiations (Chang, Cheng & Trotman, 2013).

The main mechanism that has been proposed for these differences in judgment and decision quality is that process accountability increases epistemic motivation (DeDreu et al., 2000; DeDreu et al., 2006), which leads to a more thorough evaluation of the available information (Brtek & Motowidlo, 2002; Ford & Weldon, 1981), whereas outcome accountability does not bring about similar changes in epistemic motivation (or might even detract from finding the best possible solution). Increased epistemic motivation, in turn, leads to a reduction in decision-making biases and an increase in decision quality, as long as "the normatively correct response can be identified by [...] more thorough and complex information processing" (Simonson & Nye, 1992, p.416).

Taking these findings into account, a differentiation between process and outcome accountability should also have implications for the effect of accountability on loss aversion. Specifically, we would expect that holding people process accountable should lead to a more thorough evaluation of the decision-making task, allowing them to realize that loss aversion is an irrational bias in their decision-making, and ultimately reduce it. For outcome accountability, on the other hand, we would not expect that it leads to a reduction in loss aversion. If anything, the literature so far suggests that outcome accountability might even somewhat facilitate loss aversion, because decision-makers under outcome accountability have been found to do "more of the same thing", that is, to follow their dominant strategies, which might also mean that common biases are exaggerated (e.g., Brtek & Motowidlo, 2002; Siegel-Jacobs & Yates, 1996).

Although there is no reference to outcome versus process accountability in Vieider (2009) or Pahlke et al. (2012), we are of the opinion that both of the studies predominantly manipulated process accountability. Vieider (2009) states that, for the manipulation of accountability, participants were told that they would be interviewed by an experimenter "about their choices" (Vieider, 2009, p.98) after they completed the experimental task. While this wording leaves it somewhat open for interpretation whether it is supposed to induce process or outcome accountability, it, at least in our opinion, resembles a typical process accountability manipulation much more closely than a typical outcome accountability manipulation. To provide a comparison: In their replication of the seminal Simonson and Staw (1992) study about effects of process and outcome accountability on escalating commitment, Schulz-Hardt et al.'s (2021) told participants in the process accountability condition that, after they completed working on the experimental task, a researcher would conduct an interview with them that "will refer exclusively to [their] working process" (Schulz-Hardt et al., 2021) and in which the "interviewer will only be interested in how [they] proceeded with [their] predictions and decisions, but will not care about the accuracy and the

results” (Schulz-Hardt et al., 2021). In contrast to this, participants in the outcome accountability condition were told that in the interview they would be evaluated “with regard to the accuracy of [their] predictions, and the result of [their] investment decisions” (Schulz-Hardt et al., 2021) and that “how [they] proceeded with [their] predictions and decisions will not be subject of the interview” (Schulz-Hardt et al., 2021).

Furthermore, there are three additional reasons why we believe that Vieider's manipulation predominantly would induce process accountability. One, Vieider's straight matching task does not lend itself very well to questions regarding the quality of the choices, as the task is phrased in terms of personal preferences, something that cannot be easily assessed as right or wrong. Two, because of this "unnaturalness" of the assessment of the quality of choices, we believe that if the manipulation were meant to induce outcome accountability, it would have been worded differently, i.e., by explicitly telling participants that, in the interview, they would be asked about and would have to justify the quality of their choices. Three, given that the existing literature on accountability has predominantly shown that only process accountability is effective in debiasing decisions, the fact that Vieider (2009) shows such a debiasing effect could be taken as another indication that his manipulation, for the most part, induced process accountability.

In a similar vein, even though it is never explicitly stated, it seems fairly obvious that Pahlke et al. (2012) manipulated process accountability in their experiment, as the decision-makers in the accountability condition were told that they would be interrogated by the recipient about the reasons behind their choices and, thus, would have to explain their decision-making process to the recipient.

Then again, regardless of whether Vieider (2009) and Pahlke et al. (2012) predominantly induced process accountability, outcome accountability, or a mix of the two with the manipulations they employed, they never explicitly differentiate between the two



forms of accountability and, accordingly, never specify which of the two they wanted to manipulate. Given the fact that the broader literature on the effects of accountability on judgments and decision-making has shown differential effects of process and outcome accountability, it seems imperative to follow up on the ambiguous accountability manipulations Vieider (2009) and Pahlke et al. (2012) used in their studies by cleanly differentiating the effects on loss aversion between the two forms of accountability.

### **The present research**

The initial aim of the present research was to extend Vieider's (2009) findings by testing for differential effects of process and outcome accountability on loss aversion, expecting debiasing effects for process accountability only. However, since we failed to replicate Vieider's original finding in two experiments, that is, we did not find any effects of either form of accountability on loss aversion (described below), the present study's focus shifted towards a full-fledged replication attempt. To this extent, we conducted a third experiment that focused solely on the original study's experimental conditions, employed the original materials and procedures as precisely as possible, and significantly increased the sample size compared to the original study. In the following, we will report the results of our three experiments.

### **Transparency and Openness**

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study. All data and research materials are available at our OSF repository (<https://osf.io/f7wj9/>). The designs and analysis plans for the three experiments reported here were not pre-registered.

## Experiments 1a & 1b

Our first two experiments aimed to differentiate between the effects of process and outcome accountability on loss aversion. We used experimental materials and procedures that closely resembled those of the original Vieider (2009) study. At the same time, we revised the manipulation of accountability by touching up the wording so that it clearly aimed to manipulate process accountability. Furthermore, we added a second accountability condition, namely one where we used a typical manipulation of outcome accountability, in order to test for differential effects of these two forms of accountability. Experiments 1a and 1b were, for the most part, highly similar to one another, and data collection for the two experiments happened independently in two different labs, and in close temporal proximity to one another.

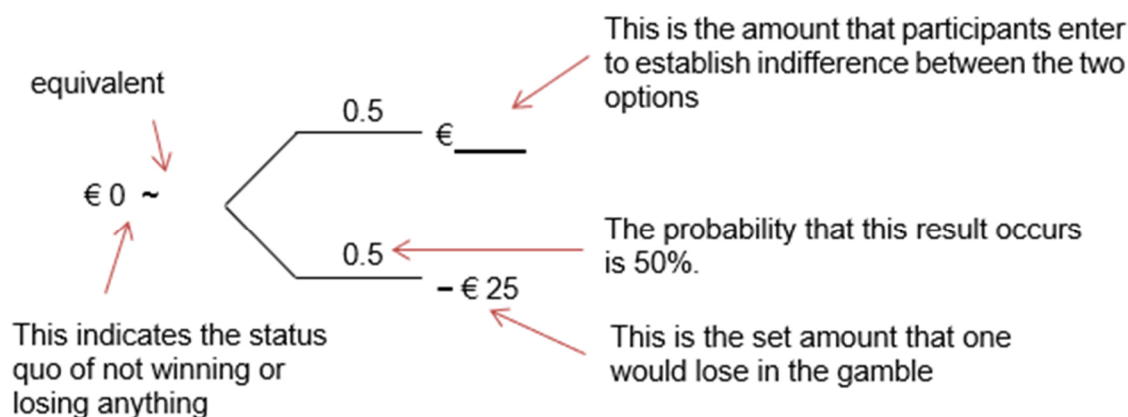
### Method experiment 1a

**Participants and design.** The sample of Experiment 1a was comprised of 102 participants (84 female, 18 male) which, for the most part (98%), were undergraduate and graduate students from the University of [location withheld to ensure double-blind review]. The rationale for this sample size was that we wanted to collect a sample that was roughly similar in size to the original Vieider (2009) study. Participants studied a diverse set of subjects, with the most prominent groups being psychology students (34%) and students enrolled in teacher training (14%). The age of the participants ranged from 19 to 24 years ( $M = 21.80$ ,  $SD = 2.51$ ). Participants were randomly assigned to one of three experimental conditions in a one-factorial between-subject design (no accountability vs. process accountability vs. outcome accountability).

**Experimental task.** Participants were presented with a German translation of the straight matching task Vieider (2009) employed, which is an adaptation of stimuli developed by Tversky and Kahnemann (1992). In this task, participants received pairs of options consisting of a set amount of 0€ (the status quo, neither winning nor losing any money) versus

a gamble with a 50% chance to lose a certain amount and a 50% chance to win a certain amount. While the amount one could lose in a specific gamble was provided in the material, participants had to fill in the other half of the gamble (the amount one could win) in such a way that they would feel indifferent between the two options of either playing the gamble or maintaining the status quo. Figure 1 depicts an exemplary trial of this straight matching task. This task allows us to approximate a loss aversion index  $\lambda$  by simply dividing the potential winnings participants entered into a specific gamble by the corresponding losses – identical to what was done in the original Vieider (2009) study. Thus,  $\lambda$  gives us the ratio of how big a gain has to be in order to be perceived as having equal value as a certain loss. Accordingly, if losses and gains are weighted equally, i.e., there is no loss aversion,  $\lambda$  takes on a value of 1. Whenever gains and losses are not weighted equally,  $\lambda$  differs from 1, with values bigger than 1 indicating loss aversion, and values smaller than 1 indicating gain seeking<sup>1</sup>. In total, the task consisted of eight outcome pairs, the first four consisted of the simple prospects described above, which were used to measure the loss attitude of our participants, and the second set of four prospects were included to check for consistency in the answers of participants, in this, we again, exactly follow the procedure of Vieider's (2009) original study.

Figure 1. Exemplary trial of the straight matching task with additional explanations.



<sup>1</sup> A recent meta-analysis by Brown, Imai, Vieider, & Camerer (2021) estimates the mean value of  $\lambda$  reported in the empirical literature on loss aversion as 1.955, with a 95% C.I. from 1.824 to 2.104. This indicates that, in general, losses are weighted about twice as heavily as (monetarily) equal gains.

**Manipulation of accountability.** In the accountability conditions, participants were informed that they would have an interview with a researcher working on the research project to which the current experiment belonged. The interview would take place right after they had completed the experimental task. This manipulation is similar to the one employed by Vieider (2009) and, in general, constitutes a typical experimental manipulation of accountability (e.g., de Langhe et al., 2011; Girandola & Gauthier, 2001; Simmons & Staw, 1992). The main difference compared to Vieider's original study is that we explicitly specified what the interview would be about. In the outcome accountability condition, participants were told that: "After this task, we want to conduct an **interview** with you. The interview will be conducted by a researcher working on the corresponding research project. In this interview, you will be asked about and will have to justify **only the results** you achieved. You are supposed to explain in the interview which amounts of money you chose. In contrast to this, it is entirely irrelevant how you proceeded while working on the task". Participants in the process accountability condition, on the other hand, were told: "After this task, we want to conduct an **interview** with you. The interview will be conducted by a researcher working on the corresponding research project. In this interview, you will be asked about and will have to justify **only the procedure** you used while working on the task. You are supposed to explain in the interview how you made your decisions. In contrast to this, it is entirely irrelevant which amounts of money you specifically chose". In both accountability conditions, participants anticipated that their interviews would be videotaped, and that some of these videos would be used for teaching and research purposes (i.e., they would be shown in seminars or at scientific congresses). Accordingly, participants had to sign an informed consent form allowing the research team to use the video material in this manner.

**Instruction and manipulation checks.** We asked participants several questions to ensure that they understood the instructions and that the manipulation of the two different forms of accountability was successful. First, participants answered a single choice item

asking what they felt accountable for while working on the task (process, outcome, or neither). Then, if they belonged to one of the two accountability conditions, participants answered a single-choice item by indicating whether the interview would be about either the process they used to complete the task or the outcome they achieved.<sup>2</sup>

**Procedure.** Participants received the first part of the experimental materials, in which they learned that, for the experimental task, they would be presented with pairs of gambles, and that they should imagine that for each pair of gambles, one gamble would be chosen for them at random and that they would have to play it afterward. They then learned that the gambles were not complete yet. Instead, in each pair one value was missing from one of the gambles (this was always the "gain value" for the gamble on the right-hand side), and their task would be to fill in the missing value in such a way that they would feel indifferent between the two gambles. Participants in the two accountability conditions were then informed that they would be interviewed after completing the task, and this interview would either exclusively be about the procedure they used to make their choices (process accountability) or the specific choices they made (outcome accountability)<sup>3</sup>.

All participants gave their informed consent and were handed the remainder of the experimental materials. Participants filled in an initial questionnaire which included the instruction and manipulation checks as well as some exploratory measures, and then started working on the straight matching task which, in total, consisted of eight prospects. The

---

<sup>2</sup> In addition to the measures for the manipulation check and the loss aversion, some exploratory measures were included. These exploratory measures were three questions that were intended to measure how much emotional strain participants experienced while working on the task. We do not further report these measures here, as they do not directly relate to our research question (in case that we had found effects of accountability, they might have helped to identify mediators for these effects). However, all additional variables can be found both in the materials and in the datasets of our experiments provided in our OSF repository (<https://osf.io/f7wj9/>).

<sup>3</sup> We are aware of the fact that this is not an idealtypical manipulation of outcome accountability. Such an idealtypical manipulation would have been achieved if the gambles would have actually been played, and the participants would then have had to justify the monetary outcomes that they received. However, doing so would have meant to alter the character of the task that Vieider (2009) had used, which did not seem feasible to us. Hence, if we wanted to stick to the original task, the present manipulation was the closest that we could get to a manipulation of outcome accountability.

experiment ended with a final questionnaire that included a suspicion check and some demographics. After completing this final questionnaire, participants were informed that there would be no interview, and that the experiment was actually over at this point. They were then fully debriefed and received 5€ for their participation.

## Results Experiment 1a

All three of our experiments were initially planned with classic frequentist hypothesis testing in mind. However, after re-focusing our studies in the direction of a replication attempt, we decided to run Bayesian analyses of our data to qualify not only the evidence for our alternative hypothesis, but also for the null hypothesis. While we will also report the results of the frequentist analyses, our conclusions will be based on our Bayesian tests. We are aware that such a mix of two different statistical approaches usually should be avoided; however, we felt it would be essential in this case to also report the frequentist results to allow the reader a direct comparison of our results with those of the original study.

All analyses were conducted with JASP (JASP Team, 2022), while data processing was performed with R (R core team, 2022) and the package *dplyr* (Wickham et al., 2022). For our Bayesian analyses, we used the standard priors implemented by JASP (e.g., Cauchy priors with  $r\text{-scale} = 0.707$  for Bayesian  $t$ -tests). Throughout our analyses, we report  $BF_{10}$  values. The  $BF_{10}$  values quantify the extent to which the data show evidence for the alternative hypothesis over the null hypothesis, with  $BF_{10}$  values larger than 1 indicating that data favor the alternative hypothesis and  $BF_{10}$  values smaller than 1 indicating that the data favor the null hypothesis. For the interpretation of the Bayes Factor, we followed Jeffreys' (1961) recommendations:  $BF_{10}$  values over 3 (under 0.33) indicate *substantial* evidence for the alternative hypothesis (null hypothesis),  $BF_{10}$  values over 10 (under 0.1) indicate *strong* evidence,  $BF_{10}$  values over 30 (under 0.033) indicate *very strong* evidence and  $BF_{10}$  values over 100 (under 0.01) are interpreted as *decisive* evidence. In a slight departure from Jeffrey's

original recommendation, we decided to modify his wording and consider  $BF_{10}$  values between 3 and  $1/3$  *inconclusive*.

**Instruction and manipulation checks.** Concerning the question what participants felt accountable for while working on the experimental task, a Bayesian contingency table test remained inconclusive, slightly favoring the alternative hypothesis that there is a difference between experimental conditions,  $BF_{10} = 1.96$  ( $\chi^2(4) = 9.74, p = .045$ , for the frequentist chi-square-test). Further testing via Bayesian multinomial tests in the individual experimental conditions showed decisive evidence that the distribution was not random for the process accountability condition,  $BF_{10} = 119.69, \chi^2(2) = 15.27, p < .001$ , as well as substantial evidence that the distribution was not random in the outcome accountability condition,  $BF_{10} = 6.60, \chi^2(2) = 8.71, p = .013$ . In contrast, evidence in the control condition was substantially in favor of the null hypothesis,  $BF_{10} = 0.33, \chi^2(2) = 3.46, p = .178$ . A closer look at the non-random distributions in the two accountability conditions showed that the majority of the participants in the process accountability condition correctly answered that they felt accountable for the process while working on their task (21 of 34 = 62%), while 9 (26%) answered that they felt accountable for the outcome, and 3 (9%) answered that they felt accountable for neither the process or the outcome. In the outcome accountability condition, a slight majority of the participants (18 of 34 = 53%) also answered that they felt accountable for the process while working on the task, whereas only 12 out of 34 (35%) answered that they felt accountable for the outcome of their decisions, and 4 (12%) answered that they had not felt accountable for either. In other words, we were able to establish process accountability, but we failed to do so for outcome accountability.

For the additional instruction check ("What will the interview following the completion of the experiment be about?"), we conducted a Bayesian contingency table test comparing the two accountability conditions and found strong evidence for a difference

between these two conditions,  $BF_{10} = 17.65$ ,  $\chi^2(1) = 8.01$ ,  $p = .005$ . Once again, we used Bayesian multinomial tests for the distribution in the individual conditions. For the process accountability condition, we found decisive evidence for differences from a random distribution,  $BF_{10} = 10584.21$ ,  $\chi^2(1) = 19.88$ ,  $p < .001$ , whereas for the outcome accountability condition we found substantial evidence that the answers of the participants were randomly distributed,  $BF_{10} = 0.31$ ,  $\chi^2(1) = 0.76$ ,  $p = .384$ . Similar to our first instruction check, process accountable participants were mostly able to correctly answer this question (30 of 34 = 88%), whereas only 14 out of 34 participants (41%) in the outcome accountability condition answered correctly.

These results indicate that our manipulation of accountability has been only partially successful. While we find rather convincing evidence that our manipulation of process accountability was working as intended, the same cannot be said for our manipulation of outcome accountability. The majority of participants in the outcome accountability condition indicated that they felt accountable for the decision-making process. Additionally, outcome accountable participants, when being asked what the interview with the researcher would be about, chose decision process and decision outcomes about equally often.

As a consequence, this experimental manipulation does not allow us to test for differential effects of process vs. outcome accountability. However, as we were able to establish the intended difference between process accountability and the control condition, it does allow us to investigate effects of process accountability on loss aversion and, thereby, test whether Vieider's (2009) findings can be replicated (as we consider his manipulation to be a manipulation of process accountability). Vieider himself did not report any manipulation or instruction checks; thus, it is unclear which psychological effects his manipulation of accountability actually had (or whether it was successful at all). Also, this makes a direct comparison of the psychological effects of our manipulation with Vieider's impossible.



Nevertheless, we have recalculated the analysis for the primary dependent variable (loss aversion) with only those participants who answered the instruction check correctly, and we found that this makes no difference to our findings. This additional analysis can be found in Supplemental Material A (<https://osf.io/f7wj9/>).

**Loss aversion.** Following Vieider (2009), we excluded 12 of the 102 participants from our analyses of loss aversion, as their answers violated stochastic dominance. Afterward, just like Vieider (2009) did in his original study, we calculated the average loss aversion of our participants by averaging the approximated loss aversion indices  $\lambda$  for each participant over the first four prospects. We then conducted a Shapiro-Wilks test to check for the normality of our data. This test rejected the assumption of normality,  $W = 0.15$ ,  $p < .001$ ; hence, similar to Vieider (2009) in his original study, we will use non-parametric tests throughout. Table 1 shows the median and mean values of the loss aversion for each of the four prospects analyzed, as well as the average loss aversion over all four prospects in the three experimental conditions.

Since there is, as of now, no comprehensive way (at least to our knowledge) to determine a Bayes Factor for a Kruskal-Wallis-Test, we only report the pairwise comparisons of our three experimental conditions in the form of Bayesian Mann-Whitney-U tests. For the comparison of the no accountability condition with the process accountability condition (the one that most closely resembles the comparison from the original Vieider study), we found substantial evidence supporting the null hypothesis,  $BF_{10} = 0.270$ ,  $W = 473$ ,  $p = .927$ ,  $r = -.015$ . The result for the comparison of non-accountable with outcome accountable participants, on the other hand, was inconclusive,  $BF_{10} = 0.502$ ,  $W = 352.5$ ,  $p = .159$ ,  $r = -.21$ . The final comparison of process accountable participants with outcome accountable participants was also inconclusive,  $BF_{10} = 0.421$ ,  $W = 496$ ,  $p = .240$ ,  $r = .18$ .

In addition, we decided to run an analysis combining the two accountability conditions. We did so for two reasons. One, as mentioned earlier, Vieider's (2009) manipulation of accountability is somewhat ambiguous, and it is unclear whether or not process accountability, outcome accountability, or a mix of both is supposed to have the strongest impact on loss aversion. Therefore, one could argue that combining the two conditions is the more appropriate test of the robustness of Vieiders (2009) findings. Two, as is obvious from the manipulation checks reported above, many of the participants in the outcome accountability felt more accountable for their decision-making process than for their decision outcomes. Thus, one might argue that both manipulations ultimately induced some sort of process accountability and, thus, should be combined. This final comparison provided substantial evidence for the null hypothesis of there being no difference in loss aversion between the no accountability condition and the combination of the two accountability conditions,  $BF_{10} = 0.266$ ,  $W = 825$ ,  $p = .389$ ,  $r = -.11$

Table 1. Median values of the loss aversion index  $\lambda$  for the four simple prospects and their average in the three experimental conditions of Experiment 1a (means in parentheses). Higher values of  $\lambda$  indicate stronger loss aversion.

	No accountability	Process accountability	Outcome accountability
$\lambda$ 1	1.64 (15.16)	1.60 (6.05)	2.00 (4.59)
$\lambda$ 2	2.00 (66.26)	2.00 (11.89)	2.50 (8.12)
$\lambda$ 3	2.00 (324.72)	2.00 (3339.58)	4.00 (1793.73)
$\lambda$ 4	2.00 (324.41)	2.00 (3339.71)	3.33 (2398.08)
Average $\lambda$	1.79 (182.64)	2.00 (1674.31)	2.85 (1051.13)

Note. The average lambda is calculated by first averaging the individual lambda values for each individual and afterwards calculating the median value per experimental condition.

## Discussion Experiment 1a

Contrary to our initial expectations, we did not find any effects of accountability (neither process nor outcome) on loss aversion. In the pairwise comparison of process accountability with no accountability, which we believe might be the apt comparison when comparing our results with those of Vieider (2009), we even found substantial evidence in favor of the null hypothesis or, in other words, evidence against an effect of process accountability. Our results, thus, contradict the previous findings reported by Vieider (2009).

However, one should interpret this finding carefully, since we deviated in certain ways from the procedure Vieider used in his study, with the most obvious difference being, of course, the change in the manipulation of accountability. Instead of using only a somewhat vague and unspecified accountability manipulation, we explicitly differentiated between process and outcome accountability, as it is now almost a standard in accountability research (see Lerner & Tetlock, 1999 for a discussion and Sharon et al., 2022 for a meta-analysis). Thus, our goal was to differentiate between the effects of process and outcome accountability on loss aversion. A goal that we were only partially able to achieve, as only our manipulation of process accountability has been successful, whereas we failed to successfully induce outcome accountability. However, this problem with the outcome accountability manipulation cannot explain the complete absence of any effects of accountability on loss aversion because, with regard to this loss aversion, even the comparison between process accountability and the control condition without accountability turned out to be nonsignificant.

Another meaningful difference between our experiment and Vieider's is that we decided to collect a heterogeneous student sample at the University [location withheld to ensure double-blind review], which consisted of students from many different backgrounds. In contrast, the sample Vieider collected in 2009 consisted mainly (94%) of students of either economics or business. Arguably, students with interest and training in economics and

business might be more familiar with the phenomenon of loss aversion as, in general, they attend some courses in mathematics and statistics. It is conceivable that this is a prerequisite for the reduction of loss aversion through accountability as, in general, accountability is assumed to reduce biases only when decision-makers are able to arrive at the correct decision by exerting higher cognitive effort (Simonson & Nye, 1992). Therefore, while students with a background in business or economics probably should be able to identify loss aversion as a bias when they think about it more closely, students without such a background might not be able to do so, as they probably are not familiar with the concept of loss aversion. It should be noted, however, that Vieider (2009) himself stated that the prospects used in his experiment were so simple that a more general subject pool should also be able to notice that loss aversion would be irrational when thinking about the task more thoroughly.

A final possible reason for our failure to replicate Vieider's (2009) findings could be that our sample size for the comparison of no accountability and process accountability in this first experiment was rather small (62 participants for the pairwise comparison of the process with no accountability condition vs. 102 participants in Vieider's 2009 study). This means it would be possible that Experiment 1a simply did not have enough test power to find the effect.

While Experiment 1b was in many ways very similar to Experiment 1a, it also addresses some of the concerns with Experiment 1a, as in Experiment 1b we collected a larger sample that also more closely resembles the composition of Vieider's (2009) sample.

## **Method Experiment 1b**

**Participants and design.** In total, 153 undergraduate and graduate students from the University of [location withheld to ensure double-blind review] participated in Experiment 1b. The rationale for the sample size of this study was that we aimed at collecting a sample in which the two conditions of primary interest (no accountability and process accountability)

combined were of similar size as Vieiders (2009) original sample. Overall, the characteristics of our sample were very similar to the sample in the original Vieider (2009) study: The average age of the participants was 23.3 years ( $SD = 4.08$ ), and 47% were male. Additionally, the sample mainly consisted of business and economics students (92%), as was the case in the original study. Participants were randomly assigned to one of three experimental conditions in a one factorial between-subject design (no accountability vs. process accountability vs. outcome accountability).

**Experimental task.** We used the same German translation of the straight matching task (adopted from Vieider, 2009) as in Experiment 1a.

**Manipulation of accountability.** The manipulation of accountability in Experiment 1b was identical to the manipulation in Experiment 1a. Participants in the two accountability conditions were informed that there would be an interview with a researcher after they finished working on the experimental task. Again, participants anticipated that their interviews would be videotaped and that some of these videos would be used for teaching and research purposes (i.e., they would be shown in seminars or at scientific congresses). The wording of the process accountability vs. outcome accountability manipulation was also similar to Experiment 1a.

**Instruction and manipulation checks.** We asked participants several questions to check whether or not they understood the instructions. In the two accountability conditions, participants answered a single-choice question regarding the content of the interview, indicating whether the interview would focus on their process during the task, their results, or both equally. Additionally, all participants were asked to indicate on an 11-point Likert scale whether, during their performance of the task, they would focus more on the outcome of their choices or on the procedure for making their choices (0 = procedure; 5 = both; 10 = outcome). Moreover, participants also had to indicate on an 11-point Likert scale (0 = not at all; 10 =

very) how likely they thought it would be that they had to justify their decisions after completing the task.

**Procedure.** The procedure of Experiment 1b was similar to that of Experiment 1a. The only differences were that Experiment 1b included some additional manipulation checks and some additional exploratory measures before and after the experimental task compared with Experiment 1a.<sup>4</sup>

## Results Experiment 1b

**Instruction and manipulation checks.** Regarding the question of what aspect of their decision-making the interview would be focused on, a Bayesian contingency table test revealed substantial evidence for differences between the two accountability conditions,  $BF_{10} = 5.40$ ,  $\chi^2(2) = 9.32$ ,  $p = .009$ , and in both conditions the observed frequencies definitively differed from a random distribution, OA:  $BF_{10} = 1173.47$ ,  $\chi^2(2) = 21.96$ ,  $p < .001$ , PA:  $BF_{10} = 131070.11$ ,  $\chi^2(1) = 25.00$ ,  $p < .001$ . Most participants in the process accountability condition classified themselves correctly (42 of 49 = 86%). In contrast, the majority of participants in the outcome accountability condition misclassified themselves by answering that the interview would be about their decision-making process (33 of 52 = 63%), and only 7 out of 52 answering correctly that the interview would be about their results (13%).

With regard to the question how likely participants felt it would be that they would have to justify themselves after the experiment, a (Bayesian) one-way ANOVA found decisive evidence for differences between the three experimental conditions,  $BF_{10} = 1696.92$ ,  $F(2, 147) = 12.35$ ,  $p < .001$ ,  $\eta_p^2 = .14$ . Post-hoc comparisons revealed that participants in both

---

<sup>4</sup> In addition to the measures for the manipulation check and the loss aversion, some exploratory measures were included. These exploratory measures were six items measuring engagement as well as three items that measured emotional strain while working on the task asked, and an additional round of manipulation and instruction checks that we asked participants after they completed the task, in order to measure how consistent the participants' answers were. We do not report these measures here, as they do not directly relate to our research question. However, all additional variables can be found both in the materials and in the datasets of our experiments provided in our OSF repository (<https://osf.io/f7wj9/>).

accountability conditions found it more likely that they would have to justify their decisions (OA:  $M = 7.28$ ,  $SD = 2.17$ , PA:  $M = 6.68$ ,  $SD = 2.41$ ) than participants in the no-accountability condition ( $M = 5.02$ ,  $SD = 2.39$ ), PA vs. no-accountability:  $BF_{10} = 28.39$ ,  $t = -3.50$ ,  $p = .002$ , OA vs. no-accountability:  $BF_{10} = 4985.09$ ,  $t = -4.85$ ,  $p < .001$ . The evidence regarding a possible difference between the two accountability conditions (which we did not expect to occur) was inconclusive,  $BF_{10} = 0.46$ ,  $t = 1.31$ ,  $p = .390$ .

For the question regarding what participants would focus on during the task, a Bayesian analysis of variance showed substantial evidence that the different experimental conditions had no impact on the participants' answers,  $BF_{10} = 0.27$ ,  $F(2, 147) = 1.64$ ,  $p = .197$ ,  $\eta_p^2 = .02$ . In all three conditions, participants (on average) gave answers close to the midpoint of the scale (no accountability:  $M = 5.22$ ,  $SD = 2.09$ ; outcome accountability:  $M = 5.56$ ,  $SD = 1.92$ ; process accountability:  $M = 4.82$ ,  $SD = 2.20$ ).

These results indicate that, similar to Experiment 1a, our manipulation of accountability again has been only partially successful. While both participants in the process and the outcome accountability condition believed it to be more likely than participants in the no accountability condition that they would have to justify themselves after the experiment, only in the process accountability condition the majority of participants were able to correctly answer what the interview would be about. Hence, our manipulation of process accountability was, for the most part, working as intended, whereas it seems that our manipulation of outcome accountability again did not have the intended effect. Thus, as in Experiment 1a, we cannot really test for differential effects of process and outcome accountability on loss aversion. However, and again similar to Experiment 1a, this still allows us to test the effect that Vieider (2009) has reported, as his manipulation of accountability mostly resembles one of process accountability (which we have successfully induced here). Notwithstanding this, we recalculated the analysis for the primary dependent variable (loss aversion) with only

those participants who answered the instruction check correctly, and we found that this did not substantially alter our results. This additional analysis can be found in Supplemental Material A (<https://osf.io/f7wj9/>).

**Loss aversion.** We followed Vieider's (2009) analytical approach closely. Of the 153 participants, 16 had to be excluded from our analyses of loss aversion, as their answers violated stochastic dominance. We then calculated the average loss aversion per participant and conducted a Shapiro-Wilks test to check for the normality of this variable; this test rejected the assumption of normality,  $W = 0.07$ ,  $p < .001$ . Hence, we will again (and similar to the original study) use non-parametric tests. The median and mean values of loss aversion for each of the four simple prospects as well as their average in the three experimental conditions of Experiment 1b can be found in Table 2. Again, we ran a series of three Bayesian Mann-Whitney-U tests, one for each pairwise comparison between the experimental conditions. For the comparison of the no accountability condition with the process accountability condition, we found substantial evidence supporting the null hypothesis that there is no difference between the two conditions,  $BF_{10} = 0.229$ ,  $W = 1031.5$ ,  $p = .868$ ,  $r = .021$ . The same was true for the comparisons of non-accountable with outcome accountable participants,  $BF_{10} = 0.230$ ,  $W = 1052.5$ ,  $p = .735$ ,  $r = .042$ , and also for the comparison of process accountable participants with outcome accountable participants,  $BF_{10} = 0.227$ ,  $W = 1056.5$ ,  $p = .717$ ,  $r = -.043$ . Based on the facts that Vieider (2009) did not differentiate or specify which form of accountability was supposed to affect loss aversion, and that a substantial part of participants in the outcome accountability indicated that they believed they would be interviewed about their decision-making process rather than their decision outcomes, we again decided to run an analysis combining the two accountability conditions. The corresponding Bayesian Mann-Whitney-U-test for differences between the no accountability condition and the combination of the two accountability conditions provided substantial evidence in favor of the null



hypothesis of no difference between the conditions,  $BF_{10} = 0.192$ ,  $W = 2084$ ,  $p = .770$ ,  $r = .03$ .

Hence, we found no evidence that any form of accountability affects loss aversion.

Table 2. Median values of the loss aversion index  $\lambda$  for the four simple prospects and their average in the three experimental conditions of Experiment 1b (means in parentheses). Higher values of  $\lambda$  indicate stronger loss aversion.

	No accountability	Process accountability	Outcome accountability
$\lambda 1$	1.20 (2.75)	1.40 (1.65)	1.50 (10.12)
$\lambda 2$	2.00 (4.27)	1.50 (1.95)	1.60 (44.28)
$\lambda 3$	2.00 (4.35)	2.00 (2.36)	2.00 (44.78)
$\lambda 4$	2.00 (4.53)	2.00 (2.34)	2.00 (73.33)
Average $\lambda$	1.75 (3.98)	1.78 (43.13)	1.54 (2.07)

Note. The average lambda is calculated by first averaging the individual lambda values for each individual and afterwards calculating the median value per experimental condition.

## Discussion Experiment 1b

Similar to Experiment 1a, the aim of Experiment 1b was to test for differential effects of process and outcome accountability on loss aversion. This time, we did so with a sample that resembled Vieider's (2009) sample with regard to the academic background of the participants, in order to maximize the comparability with the original study. Furthermore, we also increased the size of our sample to increase test power. However, we still did not replicate the effect of accountability on loss aversion originally reported by Vieider (2009). Comparing process accountability with the non-accountability condition – which is the comparison that, in our opinion, most closely resembles Vieider's manipulation of accountability – we found substantial support for the null hypothesis, that is, for the absence of effects of process accountability on loss aversion.

Although we have considerably increased our sample size compared to Experiment 1a, we still cannot exclude the possibility that, with an even larger sample, we might have found the effect of accountability on loss aversion. However, it has to be emphasized that there was not even a small trend in the predicted direction, and the same was also true for Experiment 1a. Hence, even though a type-II-error is still a possibility that we have to consider, this would, nevertheless, impose severe limitations on Vieider's earlier findings, as the effect then would have to be much weaker than previously thought.<sup>5</sup> Furthermore, in both experiments, the Bayesian tests for the particular comparison that most closely resembles the effect from the original Vieider (2009) study, namely the comparison of the process accountability condition with the no accountability condition, provided substantial evidence for the null hypothesis, lending credibility to the interpretation that there is, indeed, no such effect.

As in Experiment 1a, only the manipulation of process accountability was successful: participants believed that they would have to justify themselves after completing the experimental task and were, for the most part, able to correctly answer that the interview would be about the procedures they used to make their choices. On the other hand, outcome accountable participants once again predominantly failed to answer correctly what their interview would be about, with the majority stating it would be about their decision-making process. From this, we have to conclude that we were again not able to reliably induce outcome accountability in our participants. Although we can only speculate about it, we think the most plausible explanation is that the straight matching task that we adopted from Vieider is not particularly suitable to induce outcome accountability. As discussed earlier, the task does not lend itself to questions regarding right or wrong answers or, in general, to questions regarding the quality of choices, because it is framed more in the realms of personal

---

<sup>5</sup> In total, 153 people participated in Experiment 1b. Of these 153 participants, 90 participants were analysed in the pairwise comparison of process accountable with non-accountable participants. Thus, for the comparison of process accountability with no accountability we had about the same sample size as Vieider (2009) who analyzed a total of 102 participants regarding their loss aversion.

preferences, and the participants do not experience actual outcomes. Accordingly, it might seem much more plausible to the participants that they would be interviewed about the reasons behind their choices, even if the interview is announced to be only about the quality of their choices.

Despite these limitations regarding the outcome accountability manipulation, we believe these experiments provide sufficient evidence to raise serious questions about the robustness of the original finding by Vieider (2009), as both Experiment 1a as well as Experiment 1b showed clear evidence against any effects of (process) accountability on loss aversion. At this point, we decided it was time to take a step back and properly clarify the effect of accountability on loss aversion instead of further pursuing our original goal of differentiating the effects of process vs. outcome accountability on loss aversion. Thus, for our third experiment, we conducted a full-fledged replication attempt of the original Vieider (2009) study.

## **Experiment 2**

Experiment 2 was conducted to test the robustness of the effect of accountability on loss aversion, in other words, whether or not the effect reported by Vieider (2009) was reproducible in a direct replication. Therefore, compared with Experiments 1a and 1b, we removed the outcome accountability condition, and we changed the wording of the process accountability manipulation back so that it matched the original wording reported by Vieider (2009) as closely as possible. Apart from this, Experiment 2 largely mirrored Experiments 1a and 1b, as these, for the most part, already closely followed the procedure Vieider (2009) employed.

### **Method Experiment 2**

**Participants and design.** 200 undergraduate and graduate students from the University of [location withheld to ensure double-blind review] participated in Experiment 2,

93 of whom identified as female and 107 who identified as male. Our participants in this experiment ranged in age from 18 to 40 years ( $M = 23.2$ ,  $SD = 3.28$ ), and the vast majority of them majored in a business or economics-related field (195 of 200 = 98%). All participants were randomly assigned to one of two conditions in a one factorial between-subject design (no accountability vs. explicit accountability). The sample size was determined by an a priori power analysis using G\*Power (Faul et al., 2007), indicating that a sample of 180 participants would be sufficient to achieve 90% test-power for a medium effect size of  $d = 0.50$  (assuming  $\alpha = .05$ ). We slightly exceeded this sample size in order to make sure that we achieve the aforementioned test power even in case that some datasets might turn out to be incomplete.

**Experimental task.** We again used the same German translation of the straight matching task as in Experiments 1a and 1b, which we adopted from Vieider (2009).

**Manipulation of accountability.** We manipulated the level of explicit accountability by informing participants in the explicit accountability condition that, after they completed the experimental task, they would be asked to take their questionnaire to another room where they would be interviewed by an experimenter and that, during this interview, they would have to explain the reasons behind their decisions. In contrast, people in the no accountability condition were instructed to put their questionnaire in a cardboard box once they completed it. Furthermore, they were told that the data created by their answers would be treated as confidential, and that they would remain anonymous. This manipulation is identical to the manipulation that Vieider used in his 2009 study.

**Instruction and manipulation checks.** To check if our manipulation of accountability was successful, i.e., whether participants had understood and were aware of the instructions, we asked all participants after they completed the task to rate on an 11-point Likert scale how likely they considered it that they would have to justify their decisions (0 = not at all; 10 = very much).

We also included some of the manipulation checks from Experiments 1a and 1b to test whether participants in the explicit accountability condition felt more accountable for the process (as we would expect) or the outcome. Specifically, we asked participants after they had completed the experimental task, whether the following interview would focus on their procedure during the task or the outcome of the task, and what they had felt accountable for while working on the experimental task (the procedure, the outcome, both, neither). Additionally, we asked participants on an 11-point Likert scale on which aspect of the task they had predominantly focused on (0 = the procedure; 5 = both; 10 = the outcome).<sup>6</sup>

**Procedure.** The experimental procedure was very similar to the procedure used in Experiments 1a and 1b. After arriving at the lab, participants were welcomed and seated; then, they were handed the task description and an informed consent form, which also included the experimental manipulation. When participants had signed the informed consent, they received the task materials and a questionnaire, including the manipulation checks and some demographic questions. After completing the task and filling in the questionnaire, participants were informed that no interview would take place, and they were fully debriefed. They then signed a non-disclosure agreement and were paid 5€ for their participation.

## Results Experiment 2

**Instruction and manipulation check.** To check whether the manipulation of accountability was successful, we asked participants how likely they felt it would be that they had to justify themselves after completing the experimental task. Participants in the explicit accountability condition believed it to be much more likely that they would have to justify themselves than participants in the no accountability condition ( $M = 6.76$ ,  $SD = 2.62$  vs.  $M =$

---

<sup>6</sup> Additionally, we once more included three items for exploratory purposes that measured how much emotional strain participants were under while working on the task. We do not report these measures here, as they do not directly relate to our research question. However, all additional variables can be found both in the materials and in the datasets of our experiments provided in our OSF repository (<https://osf.io/f7wj9/>).

3.69,  $SD = 2.66$ ), with a Bayesian  $t$ -test showing decisive evidence for a difference between the two conditions,  $BF_{10} = 2.51 \times 10^{11}$ ,  $t(198) = 8.25$ ,  $p < .001$ ,  $d = 1.17$ . Hence, participants in the explicit accountability condition did expect to be held accountable more than non-accountable participants. This indicates that participants understood the experimental manipulation and were aware of it.

**Loss aversion.** Following Vieider's (2009) procedure, and similar to Experiments 1a and 1b, we excluded participants whose answers violated stochastic dominance from our analysis. This time, we had to exclude 17 of the 200 total participants (1%), which is a similar percentage of participant exclusions as in Vieider's (2009) original study, as Vieider excluded 7 of his 109 total participants (1%). We then calculated the average loss aversion for the remaining participants and conducted a Shapiro-Wilks test for normality on the average loss aversion. Like in the first two experiments, this test indicated that the assumption of normality was violated for our data ( $W = 0.30$ ,  $p < .001$ ). Hence, similar to our first two experiments and Vieider's (2009) original study, we will use non-parametric tests. Median and mean values of the average loss aversion for each of the four simple prospects, as well as their average in the two experimental conditions, can be found in Table 3.

To test for the effect of accountability on loss aversion, we calculated a Bayesian Mann-Whitney-U-test and found substantial evidence for the absence of any difference between the two experimental conditions,  $BF_{10} = 0.220$ ,  $W = 4514$ ,  $p = .316$ ,  $r = .086$ .

**Exploratory analyses.** To explore what form of accountability Vieider's (2009) manipulation actually induced, we included some exploratory manipulation checks to test what participants expected the announced interview would be about and what they retrospectively had felt accountable for.

Regarding the question what participants in the explicit accountability condition expected the interview to be about, we conducted a Bayesian multinomial test which provided

decisive evidence for differences from a random distribution,  $BF_{10} = 5.94 \times 10^{16}$ ,  $\chi^2(2) = 79.30$ ,  $p < .001$ . Most participants in the explicit accountability condition expected the interview to be about their decision-making procedure during the experimental task (75 of 105 = 71%), which is in line with our assumption that Vieider's (2009) original manipulation predominantly induces process accountability.

Furthermore, we also asked participants what, in retrospect, they had actually felt accountable for while working on the task. A Bayesian contingency table test provided very strong evidence for the  $H_0$  of no differences in the distribution of answers in the two experimental conditions,  $BF_{10} = 0.061$ ,  $\chi^2(3) = 2.91$ ,  $p = .405$ . In both conditions, a (relative) majority of our participants answered that they felt accountable for the outcome (48 of 105 = 46%, in the explicit accountability condition, and 54 of 95 = 57%, in the no accountability condition), while only a minority reported that they felt accountable for the process (28 of 105 = 27%, in the explicit accountability condition, and 19 of 95 = 20%, in the no accountability condition) or for neither the outcome nor the process (17 of 105 = 16%, in the explicit accountability condition, and 11 of 95 = 12%, in the no accountability condition).

Finally, participants answered a question about what they focused on during the task. Descriptively, non-accountable participants said they had focused slightly more on the outcome than accountable participants ( $M = 6.26$ ,  $SD = 2.22$  vs.  $M = 5.73$ ,  $SD = 2.21$ ); however, a Bayesian  $t$ -test turned out to be inconclusive,  $BF_{10} = 0.59$ ,  $t(197) = 1.69$ ,  $p = .092$ ,  $d = 0.24$ .

In conclusion, we found clear evidence that participants in Experiment 2 expected to be interviewed about their decision-making process. This indicates that the accountability manipulation employed by Vieider (2009) does, indeed, represent a manipulation of process accountability. However, we found no evidence that this manipulation actually affected what

the participants felt accountable for while working on the experimental task – at least not in retrospect.

Table 3. Median values of the loss aversion index  $\lambda$  for the four simple prospects and their average in the three experimental conditions of experiment 3 (means in parentheses). Higher values of  $\lambda$  indicate stronger loss aversion.

	No accountability	Explicit accountability
$\lambda$ 1	1.80 (2.07)	1.60 (2.36)
$\lambda$ 2	2.00 (3.85)	2.00 (2.74)
$\lambda$ 3	2.00 (5.59)	2.00 (4.26)
$\lambda$ 4	2.67 (5.77)	2.00 (4.95)
Average $\lambda$	2.00 (4.32)	2.00 (3.58)

Note. The average lambda is calculated by first averaging the individual lambda values for each individual and afterwards calculating the median value per experimental condition.

## Discussion Experiment 2

We designed Experiment 2 with the objective of testing the robustness and replicability of the previously reported effect of accountability on loss aversion (Vieider, 2009). To this end, we conducted a direct replication of the original study, using the original task, the original accountability manipulation, and a sample that very closely resembles the original sample, except for the fact that it was much larger. However, just as in our experiments 1a and 1b, we did not detect any effect of accountability on the amount of loss aversion that the participants showed. Instead, our Bayesian analysis of the data indicated that the evidence was substantially in favor of the null hypothesis, that is, it supported the interpretation that accountability does not affect loss aversion. All in all, this third experiment confirmed the results of our first two experiments, with an increased sample size and an accountability manipulation that closely resembled the original manipulation by Vieider (2009). Taking into account this close resemblance to the original study, the larger sample



size, and the results of our first two experiments, this very much questions the robustness of the original finding.

### Meta-analysis

In order to integrate the findings of our three experiments, we decided to run a meta-analysis of the effect of accountability (accountability vs. no accountability) on loss aversion for our three experiments (for Experiments 1a and 1b, we only included the comparison of the process accountability and the no accountability condition since, as already stated, Vieider's manipulation resembles process accountability more than it resembles outcome accountability). Subsequently, we reran the meta-analysis, this time including the effect Vieider found in his original 2009 study.<sup>7</sup>

Since the effect sizes for both Vieider's 2009 study and our experiments are given in the form of rank-biserial correlational coefficients, we first had to apply Fisher's z-transformation to the reported effect sizes in order to run our meta-analyses. After we transformed the correlation coefficients to Fisher z-scores and calculated the corresponding standard errors, we ran two random-effects Bayesian meta-analyses with JASP (JASP Team, 2021), employing a Cauchy prior with  $r = 0.707$ , one including only our three experiments, and one which included Vieider's 2009 study in addition to our experiments. In a final step, we then transformed the meta-analytical effect sizes back to correlation coefficients for ease of interpretation.

Integrating the findings from our three experiments reported above, the meta-analytically combined effect size was very small ( $r = 0.04$ ), and the data provided substantial

---

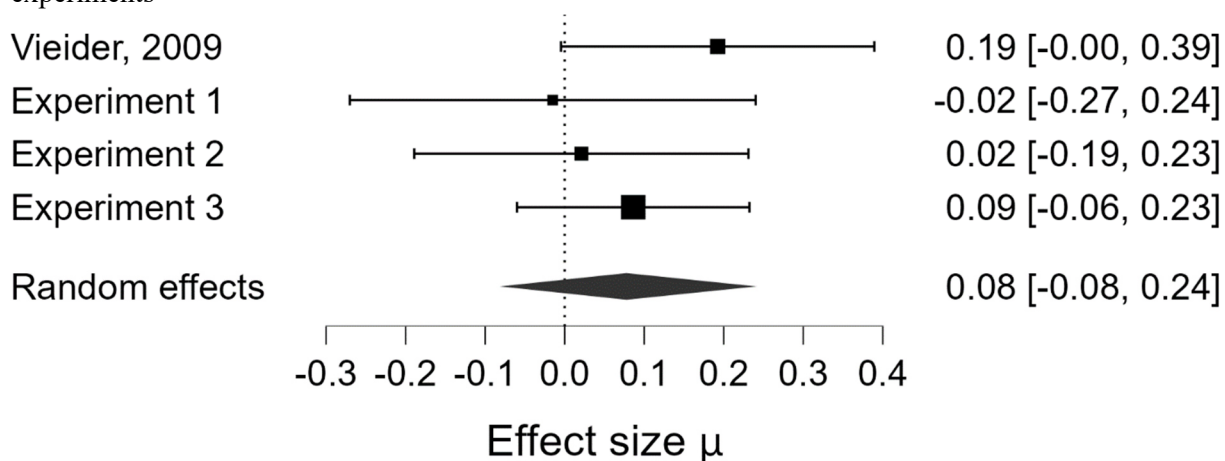
<sup>7</sup> Even though we would have liked to also include the results from Pahlke et al. (2012) in our second meta-analysis, this was not feasible because of the different approaches to measure loss aversion. While we and Vieider (2009) measured loss aversion in terms of the loss aversion index  $\lambda$ , Pahlke et al. (2012) measured loss aversion by determining the amount of safe choices in a series of mixed prospects decisions. Thus, their outcome measure is categorical while ours is continuous, which makes it very difficult and also non-informative to meta-analytically integrate the results Pahlke et al. (2012) found with Vieider's (2009) and ours.

evidence for the null hypothesis,  $BF_{10} = 0.11$ . Additionally, this meta-analysis showed substantial evidence against any between-study heterogeneity regarding the effect sizes,  $BF_{10} = 0.32$ .

When we included the original study's effect in our meta-analysis, the combined effect size remained rather small,  $r = 0.08$  (see Figure 2), and the evidence still offered substantial support for the null hypothesis,  $BF_{10} = 0.16$ . Testing again for any effect size heterogeneity between studies, we once more found that evidence substantially favored the null hypotheses of no heterogeneity,  $BF_{10} = 0.29$ .

Taken together, these two meta-analyses further substantiate our findings that the effect of accountability on loss aversion reported by Vieider (2009) is not as robust as previously thought, and that it might even be a type-I-error. Integrating the results of our three studies for the comparison of (process) accountability vs. no accountability resulted in a very small average effect size, as well as in a Bayes Factor which lent substantial support to the null hypothesis. If we take into account that the combined sample size in our three studies, even if we only consider the process accountability and the no accountability conditions in Experiments 1a and 1b, was more than three times larger than Vieider's sample in 2009 (335 vs. 102 participants) and constitutes almost 2/3 of all the available data on the effect of (process) accountability on loss aversion (All studies:  $N = 533$ ; our experiments  $N = 335$ ; Vieider, 2009 + Pahlke et al., 2012:  $N = 198$ ), we find it fair to assume that our current findings are a more reliable estimate of the true effect than those previously reported. In addition, even if we included Vieider's original study's effect in our meta-analysis, it still provided substantial support for the null hypothesis of there being no effect of accountability on loss aversion. All things considered, we feel confident to conclude that the effect of accountability, if it exists at all, is (at the very least) not very robust, and that we currently do not understand the exact conditions under which it might occur.

Figure 2. Forrest Plot of the Fisher z transformed effects observed in Vieider's (2009) and our experiments



### General discussion

The original aim of the present study was to extend the earlier findings of Vieider (2009) that holding people accountable reduced their loss aversion. Specifically, we wanted to test whether there was any difference in the effect depending on whether people were held accountable for their decision-making process or the outcome of their decisions, as it is generally accepted that these two forms of accountability have differential effects (cf. Lerner & Tetlock, 1999).

After running two experiments comparing both process and outcome accountable participants with non-accountable ones, in which we did not find any effects of either form of accountability on loss aversion, we shifted our focus towards the question of whether or not the basic finding was robust enough to be replicated at all.

Experiments 1a and 1b already provided a solid base for this replication focus, as they followed the procedure of Vieider (2009) very closely and used the same task as the original study. The design of the study, however, was slightly altered by adding a second accountability condition (outcome accountability) to allow testing for differences in the effects of different forms of accountability on loss aversion. Further, we also refined the wording of the process accountability manipulation Vieider used, to make sure that it

addresses the process (instead of the outcome) more clearly and unequivocally. Because of these deviations from the original design, Experiments 1a and 1b are conceptual rather than direct replications of Vieider (2009). Therefore, we added a direct replication attempt in our third experiment, where we tried to recreate Vieider's experiment as closely as possible, while simultaneously also increasing the sample size.

All three of our experiments failed to replicate the reduction of loss aversion in participants that were held (process) accountable, with Bayes-Factors consistently providing substantial support for the null hypotheses. This pattern of results was further substantiated when we meta-analytically combined our findings, and this outcome did not change even when we included the original effect in our meta-analysis. In conclusion, the process accountability manipulation that we employed in our experiments – and that we carried over from Vieider (2009), who, in turn, followed standard procedures used in experimental accountability research – does not seem to affect loss aversion in any meaningful way.

### **Alternative explanations**

Of course, failed replication attempts always beg the question of whether this failure is, indeed, due to an insufficient robustness of the effect, or whether there are alternative explanations for the null finding. With regard to this, there are some potential alternative explanations for our findings that have to be discussed here.

One obvious candidate for such an alternative explanation is that our experimental manipulations might not have been successful or, at least, too weak. If our manipulations did not induce (process) accountability to the same extent as the original study, it would not be surprising at all that we were unable to find the same effects as the original study. However, at least three arguments speak against this interpretation: First, our experimental manipulation of process and outcome accountability in Experiments 1a and 1b closely followed the standard procedure that has been established in this research tradition (cf. Lerner & Tetlock, 1990;

Patil et al., 2014), and our manipulation of explicit accountability in Experiment 2 followed Vieiders (2009) procedure very closely. Hence, we replicated the original study in a way that should make it unlikely for the manipulation of accountability to be significantly weaker than in the original study. Second, it has to be noted that Vieider did not report any manipulation checks in his 2009 paper, thus making it impossible to check how successful his manipulation was, and whether or not it was any more successful than ours.

Third, although not all of our manipulation checks were successful, those which were relevant with regard to replicating Vieider (2009), namely those asking how likely participants felt it to be that they would have to justify themselves after they completed the experimental task, and whether process accountable participants actually expected to be interviewed about their decision-making process, were successful: Participants in the accountability conditions felt that it would be more likely that they would have to justify themselves than participants in the no accountability condition. Additionally, the majority of participants in the process accountability condition in both Experiment 1a and Experiment 1b correctly identified what the interview following the experimental task would be about.

Further alternative explanations for the absence of any accountability effects on loss aversion could lie in possible differences between the setting and the procedures of our experiments, on the one hand, and the setting and procedures of the original study, on the other hand. It is inevitable that some slight differences between an original study and a replication study exist. For once, our studies were all conducted at German universities rather than at the Erasmus University Rotterdam in the Netherlands. Correspondingly, our experimental materials were in German rather than in English or Dutch.

Furthermore, our participants exclusively worked on the experimental task that Vieider (2009) used to measure loss aversion. In contrast to this, in Vieider's (2009) original study this straight matching task was part of a series of multiple experimental tasks that

belonged to different experiments. While we have no specific hypotheses about how this might have affected the results, we cannot exclude the possibility that this procedural difference worked in favor of the effect that Vieider (2009) reported.

Of course, it is impossible for us to exclude any of these alternative explanations with certainty, since we did not test for them and, in some cases, do not even know whether the antecedent conditions for these explanations were present or not. However, even if one or all of these alternative explanations would have caused our replication attempts to fail, one would have to conclude that the accountability effect on loss aversion is far more context-dependent and volatile than previously thought, as the original publication did not mention or imply that the effect would depend on such minuscule situational differences such as the language of the material, whether the experiment is conducted at a German or a Dutch university, or within a series of experiments rather than as a standalone experiment. Also, if it should indeed be the case that the effect is dependent on such fine-grained context variations, one would have to conclude that the effect is of very limited theoretical and practical relevance.

## **Implications**

In addition to the central implication outlined in the section above, our findings also call into question the superiority of process accountability over no accountability that has been claimed as a general effect in different domains in the literature on accountability. The results of our current study fall in line with several other studies that were recently published, and that also could not find the often-claimed positive effects of process accountability on judgment and decision quality (e.g., Chang et al., 2017; Häusser et al., 2017; Hoffmann et al., 2017). Furthermore, our current results also mirror those of some other recent replication attempts (Schulz-Hardt et al., 2021; Sluga, 2021), which failed to replicate earlier findings by Simonson and Staw (1992) and Siegel-Jacobs and Yates (1996), respectively, meaning that

the beneficial effects of process accountability reported in the original studies could not be found. Taking into account these other findings, it seems safe to say that the claim of a general beneficial effect of process accountability on judgment and decision quality is an overstatement. Rather, we have to assume that any beneficial effects of process accountability, if they exist at all, are largely context-specific, dependent on multiple situational factors, and are likely to be very small and, therefore, of very limited practical relevance (if any). This assertion is also corroborated by a new meta-analysis by Sharon, Drach-Zahavy, and Srulovici (2022), in which the authors integrated over several studies comparing process with outcome accountability. Overall, the authors found no significant difference in performance between the two forms of accountability. What they did find, however, was that outcome accountability was associated with better performance in complex tasks, and process accountability was associated with better performance in simple tasks. On top of this, even after controlling for task complexity, the heterogeneity in the subgroups remained high. These findings confirm, at least for the comparison of process and outcome accountability, how context-specific accountability effects are, given that they exist at all. Accordingly, we believe it would be important to identify further situational and contextual variables that influence the effects of (process) accountability on the quality of judgments and decisions. Besides this, however, we also think that it is important to conduct more large-scale attempts to replicate the effects of accountability on judgments and decision-making in order to clarify further which of the effects found in the existing literature are robust and which of these effects might be more context-dependent, or even non-existent.

### **Some thoughts on the manipulation of process accountability**

On top of the implications for the replicability of the Vieider (2009) effect, or effects of process accountability in general, our study also raises some doubts about the way in which researchers typically manipulate (process) accountability. Most experimental research on

accountability uses more or less the same standard manipulations in order to induce (process) accountability. Namely, it does so by announcing an interview to the participants in which they will be asked to justify their decision-making process. While rather straightforward, this manipulation is not necessarily very strong or realistic. In a real-world context, when people are held accountable, there usually are much higher stakes than performing good or bad in a simple, experimental judgment or decision-making task. Being held accountable for the investment of millions of dollars or the treatment plan of a cancer patient is very different from being held accountable for a decision on how hypothetical lotteries have to be structured in order to be attractive to the decision-maker. And more importantly, in a real-world context, there generally are consequences that one has to fear if the person, the board, or whoever one is accountable to concludes that one's decision processes or decision outcomes were not up to par, such as fines or termination of the employment contract. In contrast to this, with the typical accountability manipulation used in experimental research, the worst consequence participants would have to fear would be the unpleasant feelings one might experience when having to justify a bad decision to an expert, and this expert might then raise his or her eyebrow, or make a critical statement. Such differences beg the question of whether or not the typical experimental manipulation is strong enough to show reliable effects of process accountability, or if it can even induce process accountability properly at all. On the basis of our current findings, we have doubts about whether this manipulation is sufficiently strong: Although, according to our results, it worked in terms of making the participants aware of the interview and its content, and by making them believe that they would have to justify their decision process, it seemed to do little more than just that, as it seemed to have no impact on what participants intended to focus on during the task, or what participants subsequently reported about their focus when performing the task.

This problem is further amplified by the fact that many studies on (process) accountability do not employ any manipulation checks at all (e.g., Hoffmann et al., 2017; Kim



& Trotman, 2015; Lefebvre & Vieider, 2013; Vieider, 2009), or limit the manipulation checks to simple instruction checks, like asking participants whether or not they will be interviewed after completing the task they are working on (e.g., Brtek & Motowidlo, 2002; Chang et al., 2013; Siegel-Yacobs & Yates, 1996). Thus, in these studies, we cannot know whether or not participants actually experienced process accountability in a psychologically meaningful way.

Another issue with the research on process accountability is that there have been only very few studies that try to clarify the actual psychological mechanisms and underlying psychological effects it has on decision-makers. The main mechanism, and one of the only ones that have been proposed at all, is assumed to be that process accountability should increase epistemic motivation and lead to more rational thinking (Brtek and Motowidlo 2002; De Dreu et al., 2000). While it has been shown in a handful of studies that process accountability can increase epistemic motivation (de Langhe et al., 2011; Scholten et al., 2007; Ten Velden et al., 2010), none of these studies have properly established epistemic motivation as the mediator of the effects process accountability has on judgment and decision quality. In addition, epistemic motivation also never really gained traction in the broader accountability literature as a manipulation check for process accountability, as only a very limited number of studies ever directly measured the epistemic motivation of participants. As a result of this limited research on the psychological mechanisms behind the effects of process accountability, we think it would be a fruitful avenue for future research to further investigate the direct psychological effects and mechanisms of (process) accountability, and to identify and develop stronger manipulations of (process) accountability with higher ecological validity.

## Conclusion

Our original goal with the research we presented here was to extend and clarify the effect of (process) accountability on loss aversion reported by Vieider (2009). However, these initial plans were derailed by the results of our first two experiments. Therefore, our aim shifted towards investigating the robustness of the accountability effect Vieider (2009) had reported, thereby turning our research into a replication project. Since our three experiments and our meta-analyses found no evidence for the effect of (process) accountability on loss aversion and provided substantial evidence for the corresponding null hypothesis, we feel justified in concluding that any effect that (process) accountability might have on loss aversion is rather small, context-dependent, and thus not robust at all. We further believe that these results also question the general validity of the claim that process accountability can be used as an effective and reliable tool to reduce biases in judgments and decision-making.

## References

- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is Stronger than Good. *Review of General Psychology: Journal of Division 1 of the American Psychological Association*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Benartzi, S., & Thaler, R. H. (1995). Myopic loss aversion and the equity premium puzzle. *The Quarterly Journal of Economics*, 110(1), 73–92. <https://doi.org/10.2307/2118511>
- Brosnan, S. F., Jones, O. D., Lambeth, S. P., Mareno, M. C., Richardson, A. S., & Schapiro, S. J. (2007). Endowment effects in chimpanzees. *Current Biology: CB*, 17(19), 1704–1707. <https://doi.org/10.1016/j.cub.2007.08.059>
- Brown, A. L., Imai, T., Vieider, F., & Camerer, C. (2021). Meta-Analysis of Empirical Estimates of Loss-Aversion. In *Available at SSRN*. <https://doi.org/10.2139/ssrn.3772089>
- Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *The Journal of Applied Psychology*, 87(1), 185–191. <https://doi.org/10.1037/0021-9010.87.1.185>
- Camerer, C. (2005). Three Cheers—Psychological, Theoretical, Empirical—for Loss Aversion. *JMR, Journal of Marketing Research*, 42(2), 129–133. <https://doi.org/10.1509/jmkr.42.2.129.6228>
- Chang, L. J., Cheng, M. M., & Trotman, K. T. (2013). The effect of outcome and process accountability on customer–supplier negotiations. *Accounting, Organizations and Society*, 38(2), 93–107. <https://doi.org/10.1016/j.aos.2012.12.002>
- Chang, W., Atanasov, P., Patil, S., Mellers, B. A., & Tetlock, P. E. (2017). Accountability and adaptive performance under uncertainty: A long-term view. *Judgment and Decision Making*, 12(6), 610–626.
- De Dreu, C. K. W., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement. *Journal of Personality and Social Psychology*, 90(6), 927–943. <https://doi.org/10.1037/0022-3514.90.6.927>

- De Dreu, C. K., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: a motivated information-processing approach to integrative negotiation. *Journal of Personality and Social Psychology*, *79*(6), 975–987. <https://doi.org/10.1037/0022-3514.79.6.975>
- De Dreu, C. K. W., & van Knippenberg, D. (2005). The possessive self as a barrier to conflict resolution: effects of mere ownership, process accountability, and self-concept clarity on competitive cognitions and behavior. *Journal of Personality and Social Psychology*, *89*(3), 345–357. <https://doi.org/10.1037/0022-3514.89.3.345>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191, <https://doi.org/10.3758/BF03193146>
- Ford, J. K., & Weldon, E. (1981). Forewarning and accountability: Effects on memory-based interpersonal judgments. *Personality & Social Psychology Bulletin*, *7*(2), 264–268. <https://doi.org/10.1177/014616728172013>
- Harbaugh, W. T., Krause, K., & Vesterlund, L. (2001). Are adults better behaved than children? Age, experience, and the endowment effect. *Economics letters*, *70*(2), 175–181. [https://doi.org/10.1016/S0165-1765\(00\)00359-1](https://doi.org/10.1016/S0165-1765(00)00359-1)
- Häusser, J. A., Frisch, J. U., Wanzel, S., & Schulz-Hardt, S. (2017). Effects of Process and Outcome Accountability on Idea Generation. *Experimental Psychology*, *64*(4), 262–272. <https://doi.org/10.1027/1618-3169/a000368>
- Hoffmann, J. A., Gaissmaier, W., & von Helversen, B. (2017). Justifying the judgment process affects neither judgment accuracy, nor strategy use. *Judgment and Decision Making*, *12*(6), 627–641.
- JASP Team (2022). JASP (Version 0.16.2) [Computer software].
- Jeffreys, H. (1961). *The Theory of Probability*. OUP Oxford.
- Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, *5*(1), 193–206. doi: 10.1257/jep.5.1.193

- Kanngiesser, P., Santos, L. R., Hood, B. M., & Call, J. (2011). The limits of endowment effects in great apes (*Pan paniscus*, *Pan troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*). *Journal of Comparative Psychology*, *125*(4), 436–445.  
<https://doi.org/10.1037/a0024516>
- Kim, S., & Trotman, K. T. (2015). The comparative effect of process and outcome accountability in enhancing professional scepticism. *Accounting and Finance*, *55*(4), 1015–1040. <https://doi.org/10.1111/acfi.12084>
- Köbberling, V., & Wakker, P. P. (2005). An index of loss aversion. *Journal of economic theory*, *122*(1), 119–131. <https://doi.org/10.1016/j.jet.2004.03.009>
- Lakshminaryanan, V., Chen, M. K., & Santos, L. R. (2008). Endowment effect in capuchin monkeys. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *363*(1511), 3837–3844. <https://doi.org/10.1098/rstb.2008.0149>
- Langhe, B. de, van Osselaer, S. M. J., & Wierenga, B. (2011). The effects of process and outcome accountability on judgment process and performance. *Organizational Behavior and Human Decision Processes*, *115*(2), 238–252.  
<https://doi.org/10.1016/j.obhdp.2011.02.003>
- Lefebvre, M., & Vieider, F. M. (2013). Reining in excessive risk-taking by executives: the effect of accountability. *Theory and Decision*, *75*(4), 497–517.  
<https://doi.org/10.1007/s11238-012-9335-2>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- List, J. A. (2004). Neoclassical theory versus prospect theory: Evidence from the marketplace. *Econometrica: Journal of the Econometric Society*, *72*(2), 615–625.  
<https://doi.org/10.1111/j.1468-0262.2004.00502.x>
- Maddux, W. W., Yang, H., Falk, C., Adam, H., Adair, W., Endo, Y., Carmon, Z., & Heine, S. J. (2010). For whom is parting with possessions more painful? Cultural differences in the endowment effect. *Psychological Science*, *21*(12), 1910–1917.  
<https://doi.org/10.1177/0956797610388818>
- Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance*, *53*(5), 1775–1798. <https://doi.org/10.1111/0022-1082.00072>

- Pahlke, J., Strasser, S., & Vieider, F. M. (2012). Risk-taking for others under accountability. *Economics Letters*, *114*(1), 102–105. <https://doi.org/10.1016/j.econlet.2011.09.037>
- Patil, S. V., Vieider, F., & Tetlock, P. E. (2014). Process versus outcome accountability. *The Oxford Handbook of Public Accountability*, 69–89. doi: 10.1093/oxfordhb/9780199641253.013.0002
- Putler, D. S. (1992). Incorporating Reference Price Effects into a Theory of Consumer Choice. *Marketing Science*, *11*(3), 287–309. <https://doi.org/10.1287/mksc.11.3.287>
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. [computer software]
- Rozin, P., & Royzman, E. B. (2001). Negativity Bias, Negativity Dominance, and Contagion. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *5*(4), 296–320. [https://doi.org/10.1207/S15327957PSPR0504\\_2](https://doi.org/10.1207/S15327957PSPR0504_2)
- Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. <https://doi.org/10.1007/BF00055564>
- Scholten, L., van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. W. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, *43*(4), 539–552. <https://doi.org/10.1016/j.jesp.2006.05.010>
- Schulz-Hardt, S., Rollwage, J., Wanzel, S. K., Frisch, J. U., & Häusser, J. A. (2020). Effects of process and outcome accountability on escalating commitment: A two-study replication. *Journal of Experimental Psychology: Applied* *27*(1), 112–124. <https://doi.org/10.1037/xap0000321>
- Sharon, I., Drach-Zahavy, A., & Srulovici, E. (2022). The Effect of Outcome vs. Process Accountability-Focus on Performance: A Meta-Analysis. *Frontiers in Psychology*, *13*. <https://doi.org/10.3389/fpsyg.2022.795117>
- Shefrin, H., & Statman, M. (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, *40*(3), 777–790. <https://doi.org/10.1111/j.1540-6261.1985.tb05002.x>

- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of Procedural and Outcome Accountability on Judgment Quality. *Organizational Behavior and Human Decision Processes*, 65(1), 1–17. <https://doi.org/10.1006/obhd.1996.0001>
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, 51(3), 416–446. [https://doi.org/10.1016/0749-5978\(92\)90020-8](https://doi.org/10.1016/0749-5978(92)90020-8)
- 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>
- Sluga, A. (2021). *Hour of Judgment: On judgment, decision making, and problem solving under accountability* (Nr. EPS-2021-520-F&A). <https://repub.eur.nl/pub/136967/hffinaluploadissertation.pdf>
- Ten Velden, F. S., Beersma, B., & De Dreu, C. K. W. (2010). It takes one to tango: the effects of dyads' epistemic motivation composition in negotiation. *Personality & Social Psychology Bulletin*, 36(11), 1454–1466. <https://doi.org/10.1177/0146167210383698>
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39–60. [https://doi.org/10.1016/0167-2681\(80\)90051-7](https://doi.org/10.1016/0167-2681(80)90051-7)
- Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics*, 112(2), 647–661. <https://doi.org/10.1162/003355397555226>
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061. <https://doi.org/10.2307/2937956>
- Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323. <https://doi.org/10.1007/BF00122574>
- Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological Bulletin*, 134(3), 383–403. doi: 10.1037/0033-2909.134.3.383

- Vieider, F. M. (2009). The effect of accountability on loss aversion. *Acta Psychologica*, 132(1), 96–101. <https://doi.org/10.1016/j.actpsy.2009.05.006>
- Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: an experimental analysis. *Journal of Economic Behavior & Organization*, 33(2), 167–184. [https://doi.org/10.1016/S0167-2681\(97\)00089-9](https://doi.org/10.1016/S0167-2681(97)00089-9)
- Wickham H, François R, Henry L, Müller K (2022). *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>.



## Appendix B

### Manuscript 2

What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms.

**What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms.**

**Johannes Rollwage, Christian Treffenstädt & Stefan Schulz-Hardt**

University of Goettingen

## Abstract

Holding people accountable for their decision-making process (process accountability) instead of the outcomes of their decisions (outcome accountability) has been heralded as a valuable tool for debiasing judgments and decisions. However, recent research has challenged the assumed superiority of process accountability, as a couple of studies failed to replicate earlier findings of process accountability facilitating (and outcome accountability lowering) judgment and decision quality. In this paper, we discuss potential methodological shortcomings of the previous research, focusing on the most commonly used experimental manipulation of accountability, namely announcing an interview to study participants in which they have to justify themselves for their process or their outcomes. Subsequently, we test whether this interview-manipulation reliably produces the expected psychological effects. Two experiments ( $n = 236$  and  $n = 68$ ) show that the experimental manipulation is only partially understood by and salient to participants. Despite this, the interview-manipulation is mostly successful in inducing the psychological states of felt process and outcome accountability respectively. However, it does so without significantly affecting either of the two psychological mechanisms (epistemic motivation and stress) which, based on the existing literature, supposedly mediate the effects of process and outcome accountability, respectively. These findings provide evidence that the usual experimental manipulation of process and outcome accountability does not reliably trigger any of the previously proposed psychological mechanisms.

*Keywords:* process accountability, outcome accountability, epistemic motivation, stress, psychological mechanism

## **What does the typical experimental manipulation of process vs. outcome accountability actually manipulate? A comprehensive investigation of possible psychological mechanisms.**

Human judgment and decision-making are often found to be irrational and clouded by many different forms of cognitive bias, such as the confirmation bias (Nickerson, 1998), the sunk cost fallacy (Arkes & Blumer, 1985; Thaler, 1980), the decoy effect (Huber et al., 1982), or loss aversion (Tversky & Kahnemann, 1991), to name just a few. Because of this, researchers and practitioners are always searching for ways to debias and improve judgment and decision-making. One structural feature that has been touted as a way to achieve this goal is *accountability*. Accountability can be defined as the pressure or expectation of having to justify one's decisions to some kind of authority (Lerner & Tetlock, 1999; Simonson & Nye, 1992). In general, decision-makers are often held accountable for their decisions in one form or another, as accountability is essential to social and organizational relationships. Without accountability, bad decisions would have little or no consequences. Thus, accountability is a structural feature that is supposed to make judgments and decisions more thorough and less biased.

However, even though research in the past sometimes did claim a general benefit of holding people accountable (Tetlock, 1983; Tetlock and Boettger, 1989), accountability usually is not understood as a unitarian concept. Instead, there are multiple ways in which accountability can be implemented, and these different forms of accountability have been shown to affect judgment and decision quality differently from one another.

### **Process versus outcome accountability**

Two of the most important forms of accountability are outcome and *process accountability* (for an overview, see Patil et al., 2014). Outcome accountability is

characterized as a form of accountability in which a person has to justify or is evaluated based on their results. It does not matter how they reached their decisions or achieved certain results, but only whether the outcomes of these decisions are good or bad. This form of accountability is very prevalent in the business world. For example, a CEO will mainly be held accountable for the net results of their company, (mostly) regardless of how those results were achieved.

On the other hand, process accountability is a form of accountability in which a person has to justify or is evaluated based on how they arrived at their judgments or decisions. In other words, if one is process-accountable, it only matters whether or not one has employed a rational or defensible decision-making process. On the contrary, the consequences of the decision, for example, whether the hired candidate turned out to be successful or not in her job, do not matter. This form of accountability is most common within civil services or the public health care sector: For example, an excise officer will be evaluated based on following the correct process when reviewing tax filings, not based on how much or how little tax returns are being allotted based on their reviews.

Experimental research has led to the conclusion that process accountability seems to be a more promising way to implement accountability, as it has been reported to be associated with increased judgment and decision quality in various tasks. For example, process accountability has been reported to improve interview validity (Brtek & Motowidlo, 2002), help to overcome the fixed-pie bias in negotiations (de Dreu et al., 2000), and improve consistency and calibration in judgment tasks (Siegel-Jacobs & Yates, 1996). In contrast, beneficial effects of outcome accountability have been mostly absent from the research literature. In some cases, outcome accountability has even been found to have even detrimental effects on performance (e.g., Arkes et al., 1986; Siegel-Jacobs & Yates, 1996).

To explain these differential effects of process and outcome accountability, two different psychological mechanisms have been proposed in the literature so far.

Beneficial effects of process accountability have primarily been attributed to an increase in epistemic motivation as a consequence of process accountability (Brtek & Motowidlo, 2002; DeDreu et al., 2006; de Langhe et al., 2011). Epistemic motivation describes the desire to achieve an accurate understanding of the world or, more specifically, of the task at hand and is strongly associated with the willingness to engage in systematic and effortful information processing (DeDreu et al., 2006; Kruglanski, 1989). While previous research has shown that process accountability can lead to an increase in epistemic motivation, no such connection seems to exist between outcome accountability and epistemic motivation (DeDreu et al., 2000; DeDreu et al., 2006; de Langhe et al., 2011).

Outcome accountability, on the other hand, has been associated with an increase in stress to super-optimal levels (Siegel-Jacobs & Yates, 1996). High stress levels, in turn, have been found to have detrimental effects on the judgment and decision-making process (Keinan et al., 1987; Yates, 1990). In contrast to this, it has been theorized that process accountability leads to moderate increases in stress levels (Siegel-Jacobs & Yates, 1996), which have been found to be beneficial to the quality of judgments and decisions (Janis & Mann, 1977; Yates, 1990).

As mentioned above, it is a long-held belief in research on accountability that process accountability, in general, has beneficial effects and is superior to holding people outcome-accountable, or not accountable at all. Consequently, this notion has found entry into influential review articles and textbooks (e.g., Bovens et al., 2014; Lerner & Tetlock, 1999; Schneider & Shanteau, 2003). However, recently more and more studies have shown up that fail to support, or even contradict, this view. For example, some studies could not detect any beneficial effects of process accountability on idea generation (Häusser et al., 2017), judgment accuracy (Hoffmann et al., 2017), or forecasting accuracy (Chang et al., 2017). In addition, there have been a couple of studies (Schulz-Hardt et al., 2021; Rollwage et al., 2022)

that also question the robustness of the previously reported debiasing effects of process accountability, as they were unable to replicate the earlier findings of Simonson and Staw (1992) and Vieider (2009), respectively. A recent meta-analysis by Sharon et al. (2022) further corroborates these findings, as it reports that neither form of accountability is generally superior for improving decision quality. Rather, there are situational characteristics, in this case, task complexity, that affect which form of accountability is more beneficial. Furthermore, there are multiple, as of now unpublished, studies that question other earlier findings on the benefits of process accountability. For example, in his PhD-thesis, Sluga (2021) found that the findings by Siegel-Jacobs and Yates (1996) did not replicate. Also, a series of experiments from the DFG-project "Prozess- und Ergebnisverantwortlichkeit bei Gruppenentscheidungen" (Häusser & Schulz-Hardt, 2017) were unable to replicate findings on the effects of process and outcome accountability by De Dreu et al. (2000), De Langhe et al. (2011), Girandola & Gauthier (2001) and Scholten et al. (2007). Hence, it seems time to investigate what causes this increasing number of null findings and the problems with replicating earlier findings regarding process and outcome accountability.

A possible problem that could explain the recent null findings and failed replication attempts is that the most commonly used experimental manipulation of accountability might not be as effective as previously thought. This experimental manipulation has been used in practically all of the studies that later failed to replicate (e.g., Girandola & Gauthier, 2001; Simonson & Staw, 1992; Vieider, 2009) and is, in general, a very typical manipulation of accountability (e.g., Brtek & Motowidlo, 2002; Girandola & Gauthier, 2001; Simonson & Staw, 1992). This very common accountability manipulation consists of announcing an

interview to study participants, and the specific focus of this interview is used to induce either process or outcome accountability.<sup>1</sup>

One fairly representative example of such an interview-manipulation of the two types of accountability can be found in a study by Verwaeren and Nijstad (2022), in which participants had to maximize the profits of a virtual lemonade stand. In this study, the authors presented participants in the process accountability condition with the following instruction before working on the task: “The researcher will ask you to justify the thinking process or strategy that you followed during the task. It will not be about the profits that you made, just about the strategy you followed” (Verwaeren & Nijstad, 2022, p. 428). In contrast, the instruction given to participants in the outcome accountability condition was as follows: “The researcher will ask you to justify your result, the total profit that you obtained over the 20 rounds. It will not be about how you approached the task, just about the result” (Verwaeren & Nijstad, 2022, p. 428).

As can be seen from these examples, in the case of process accountability, the interview usually is announced to focus on the reasoning, the procedures, and/or the strategies that participants used when they made their decisions in the experimental task at hand and which they have to justify to an expert. In contrast, when the manipulation is used to induce outcome accountability, the interview is usually announced to focus solely on results, which means on the quality or accuracy of the decisions the participants made.

While this interview-manipulation looks very similar in a great variety of studies, some small variations exist. For example, sometimes the interview is announced to be conducted right after the completion of the experimental task (e.g., de Langhe et al., 2011;

---

<sup>1</sup> This interview manipulation is by far the most common manipulation of process accountability found in the literature. While there are alternative manipulations of outcome accountability, especially in experiments that directly compare process and outcome accountability, it is common to also manipulate outcome accountability via the interview (e.g., Brtek & Motowidlo, 2002; Girandola & Gauthier, 2001; Simonson & Staw, 1992).



Siegel-Jacobs & Yates, 1996; Vieider, 2009); other times, it is announced to be conducted at a later date (e.g., De Dreu et al., 2000; Girandola & Gauthier, 2001; Simonson & Staw, 1992). Further variations include either actually conducting the interview (e.g., Siegel-Jacobs & Yates, 1996) or not conducting the interview (e.g., Simonson & Staw, 1992), as well as either providing exemplary interview questions (e.g., Häusser, et al., 2017) or not providing such exemplary questions (de Dreu & van Knippenberg, 2005). However, at its core, all these studies use the same experimental manipulation.

While this interview-manipulation is relatively straightforward to implement, it might be argued that it does not adequately capture what being held accountable entails in the real world. When people are held accountable in the laboratory, the decisions and judgments they are accountable for are usually of little to no importance to them. Generally, these decisions are either hypothetical (e.g., studies on escalating commitment or loss aversion) or inconsequential (e.g., multi-cue judgment tasks). In addition, the way in which study participants are held accountable entail such little consequences for suboptimal decisions that they are almost negligible. As described above, participants in laboratory studies are primarily held accountable by being told that an expert will interview them and that they will have to justify either their decision-making process or their decision outcomes to this expert. Based on this, the most severe consequence participants in laboratory experiments on accountability have to fear is the somewhat unpleasant feelings they might experience when they have to justify their suboptimal decisions to an expert who might react negatively to this explanation. Given this weakness of the accountability manipulation in the laboratory, it would not be inconceivable that it might not be strong enough to reliably produce (differential) effects of the two types of accountability.

This uncertainty regarding the effectiveness of the typical experimental manipulation of process accountability could be easily clarified if the respective studies had consistently

reported measures showing the effectiveness (or ineffectiveness) of the experimental manipulations and their impact (or lack of impact) on the psychological measures that are assumed to lead to differential effects of judgment and decision quality. However, experimental studies on accountability often either include only minimal manipulation checks, like simple instruction checks (e.g., Brtek & Motowidlo, 2002; Chang et al., 2013; Siegel-Yacobs & Yates, 1996), or no manipulation checks at all (e.g., Hoffmann et al., 2017; Kim & Trotman, 2015; Lefebvre & Vieider, 2013; Vieider, 2009) – let alone measures of the intended psychological mediators of accountability effects. Because of this, for many studies, it is unclear what kind of effect the manipulation of process and outcome accountability had, and to what extent these manipulations were successful.

### **The present research**

Given the earlier mentioned failed replications, our concerns about the strength of the interview-manipulation of accountability, and the often missing manipulation checks in the literature, we think it is crucial to systematically test how participants perceive this very common experimental manipulation of process and outcome accountability, and what kinds of psychological effects it produces. To do so, we will check the effectiveness of the manipulation in a three-step procedure, as there are three crucial points at which this manipulation (or any manipulation) could fail. The first critical point is whether the manipulation is understood by participants and salient to them while working on an experimental task. Obviously, if participants do not understand the manipulation or are not aware of it, then there is no way for the accountability manipulation to exert any further psychological effects. The second critical step, given that participants correctly understood the manipulation, is that the manipulation has to induce the psychological state it is supposed to induce; in this case this would mean that the manipulation increases the felt (process or outcome) accountability of participants. Finally, the third step required for a successful

manipulation is to trigger the psychological mechanisms that are supposed to lead to gains (or losses) in judgment and decision quality. In other words, according to the literature, the manipulation should either affect participants' epistemic motivation, or alter their stress levels, or both. If the process (or outcome) accountability manipulation fails at any of these three steps, it is unlikely that it will affect judgment and decision quality in line with the theoretical predictions.

## **Hypotheses**

Based on the three steps needed for a successful manipulation described above and the findings reported regarding process and outcome accountability in previous research, we formulated a series of hypotheses regarding the effects of the most common accountability manipulation. To test these hypotheses, we designed a one-factorial experiment with three levels in which we compared no accountability with process accountability and outcome accountability.

For the first step of a successful manipulation, described above and in line with previous findings, we assume that participants to whom an interview about their decision-making process or decision outcomes is announced are aware of the content of the interview. Furthermore, it is assumed that participants in both accountability conditions also believe it to be more like that they will have to justify themselves subsequently to working on an experimental task.

*Hypothesis 1a:* Participants in the process accountability condition will answer that they will be interviewed about their decision-making process when asked.

*Hypothesis 1b:* Participants in the outcome accountability condition will answer that they will be interviewed about their decision outcomes when asked.

*Hypothesis 2a:* Participants in the process accountability condition will state a higher likelihood of having to justify themselves than participants in the control condition.

*Hypothesis 2b:* Participants in the outcome accountability condition will state a higher likelihood of having to justify themselves than participants in the control condition.

Regarding the second step described above, namely whether the manipulation induces the psychological state it aims to induce, we predict that the announcement of interviews about the decision-making process and the decision outcomes will lead to increased feelings of felt process and outcome accountability, respectively.

*Hypothesis 3:* Participants in the outcome accountability condition report higher levels of felt outcome accountability than participants in the process accountability and the control condition.

*Hypothesis 4:* Participants in the process accountability condition report higher levels of felt process accountability than participants in the outcome accountability and the control condition.

Finally, for the third hurdle a successful manipulation has to take, one would expect that process accountability would cause an increase in epistemic motivation, based on the previous accountability literature. At the same time, one would not anticipate finding similar effects for outcome accountability. Furthermore, or alternatively, both forms of accountability should produce stress, but outcome accountability should lead to higher stress levels than process accountability.

*Hypothesis 5:* Participants in the process accountability condition show higher epistemic motivation than participants in the outcome accountability and the control condition.

*Hypothesis 6:* Stress levels will increase more from a baseline, pre-manipulation, stress measurement to a critical, post-manipulation, stress measurement for participants in the two accountability conditions than for participants in the control condition.

*Hypothesis 7:* Stress levels will increase more from a baseline, pre-manipulation, stress measurement to a critical, post-manipulation, stress measurement for participants in the outcome accountability condition than for participants in the process accountability condition.

## **Overview of Experiments**

To test the hypotheses described above, we conducted two experiments. In the first experiment, we aimed for maximum test power by collecting a large sample. To that end, we decided to conduct Experiment 1 as an online study, as this allowed us to collect a much larger sample than would be possible in a laboratory study, given that it was conducted during the height of the Covid-19 pandemic, and we felt it would have been irresponsible to conduct a large laboratory study given the potential risk of spreading the infection by forcing participants to partake in the experiment in close proximity to one another. However, since online studies have been seldomly used when researching the effects of accountability, we conducted a second experiment in our laboratory – with a comparatively smaller sample – to show that the results from our online study were reliable and comparable to those found in the laboratory. Otherwise, Experiment 2 was mostly identical to Experiment 1.

## **Experiment 1**

### **Method**

#### **Participants and Design.**

We recruited 241 participants (155 female, 82 male, two other, and two who did not provide an answer), ranging in age from 19 years to 64 years ( $M = 26.82$ ,  $SD = 6.50$ ), for an online experiment using multiple platforms such Facebook, SurveyCircle, and Instagram.

Participants were randomly assigned to one of three experimental conditions in a one-factorial between-subject design (no accountability vs. process accountability vs. outcome accountability). A priori, we computed a power analysis using G\*Power (Faul et al., 2007). Assuming an effect of a medium size ( $f = 0.25$ ), with  $\alpha = 0.05$  and a power of .95, the required sample size was  $n = 252$ . Our final sample was  $n = 236$  (because time constraints kept us from collecting any more data, and 5 participants had to be excluded from our analyses as they indicated that they did not work seriously on the experimental task)<sup>2</sup>, leaving us with a power of .94 for our main analyses. We felt it was paramount to have such a high test power for this study, as we wanted to draw clear conclusions even if we found null results regarding the effects of the typically employed accountability manipulations. Accordingly, we planned for a sample size that reduced the type II error rate to a level close to our type I error rate.

Participants were compensated with the chance to participate in a cash raffle of 10 x 20€ and 10 x 10€, or by earning partial course credit.

### **Procedure and Materials.**

The online experiment was constructed using the software *alfred3* (Treffenstädt et al., 2021). After opening the link to the online experiment, participants were presented with a welcome text describing the experiment's procedure and the incentive structure. They then read an informed consent form to which they had to agree to participate in the experiment. Afterward, they were screened for past participation in the experiment and then filled in the state anxiety subscale of the German version of the State-trait-anxiety-inventory (STAI; Spielberger et al., 1970; Laux et al., 1981) to measure their baseline stress level (for more details on the dependent variables, see below). Then followed the manipulation of

---

<sup>2</sup> The time constraints mentioned here also had another side effect, as they caused a slight imbalance in the sample sizes of our three experimental conditions. The reason for this is that participants in the two accountability conditions discontinued the experiment at a higher rate than participants in the no accountability condition, and we were unable to equalize this imbalance in time.

accountability and some simple instruction checks, after which participants began to work on a multi-cue estimation task. Subsequently, participants filled in additional questionnaires (i.e., manipulation checks, the second measurement of stress, epistemic motivation, systematic information processing, a suspicion check<sup>3</sup>, some demographics, and a single item asking whether participants worked seriously on the experiment). Then they received feedback on the accuracy of their estimates. Finally, they were thanked for their participation and fully debriefed.

### **Manipulation of accountability**

In order to manipulate the two forms of accountability, we announced an interview in which participants would be asked to justify themselves, either regarding the outcome they achieved in the experimental task or the process they used while working on the task.

In all conditions, participants were initially informed that some of them would be randomly chosen to participate in an interview with a researcher and an expert for the task. Later in the experiment, participants in the two accountability conditions were presented with a screen that announced to them that they had been chosen for an interview that would last approximately 15 minutes and would be conducted at a separate time via the Videoconference System of the University of Göttingen, while participants in the no accountability condition were not presented with such an announcement. Participants in the two accountability conditions then had to enter two dates and times in the next two weeks at which they would be available for the interview, as well as their e-mail address, to schedule the interview and receive the link to the video call.

---

<sup>3</sup> While inspecting the answers participants gave on the suspicion check item, we found a total of 12 answers that were, to some extent, peculiar. None of these specifically mentioned any of our hypotheses, however most of these did mention the interview and/ or the different psychological measures we employed in some form as the main focus of our experiment. Thus, we reran all of our analysis excluding the 12 participants with peculiar suspicion check answers. However, this exclusion did not change our results in any meaningful way, which is why we retained all participants for the results reported here. The additional analyses with the exclusion can be found in Supplement A

Depending on which accountability condition participants were assigned to, the announcement either said they would be interviewed about the outcomes they achieved during the task, i.e., how accurate their estimates were, or about the process of completing the task, i.e., what kind of strategies they used to come up with their specific estimates. More specifically, participants in the outcome accountability condition were told that: “In this interview, you will be evaluated based on your **results**, i.e., the accuracy of your estimates. For this, only your **accuracy** matters; the interviewers are only interested in how good your estimates are. Your course of action, i.e., the procedures you used to come with your estimates, does not matter”. In contrast to this, participants in the process accountability condition were told: “In this interview, you will be evaluated based on the **strategy** you used for your estimates. During this evaluation of your **work process**, you will be asked to justify how you came up with your estimates. For this, only your **procedure** for making estimates matters; the interviewers are only interested in how you came up with your estimates. The results, i.e., how accurate your estimates were, do not matter”.

### **Multi-cue estimation task**

After the interview had been announced and the participants had answered some short manipulation check items, participants began working on a multi-cue estimation task. In this task, participants had to estimate the price per person in Euros of ten different package holiday offers. Each offer was presented with a picture and a short text including information regarding the location and the length of the holiday, the airline used for transit, the star rating of the hotel, and information about the size and configuration of the particular hotel room as well as information regarding the catering. The ten offers were identical for all participants; however, their order was randomized. The stimuli (i.e., the holiday offers) were developed and selected on the basis of a separate pretest. For this pretest, a total of 45 stimuli were identified by searching on an online travel portal (weg.de) that included all of the needed cues and a price per person (which was then taken as the true value). The stimuli were then



processed by combining the relevant cues into standardized text descriptions of the holiday offers. Ultimately, we choose 10 of the 45 stimuli at random for this online-experiment, as there was no discernible difference between the stimuli.

We used a multi-cue estimation task for this experiment, because (multi-cue) judgment and estimation tasks have been a typical type of task in previous accountability research (e.g., Siegel-Jacobs & Yates, 1996; de Langhe et al., 2011). Thereby, choosing this task allowed us to provide a conceptual replication of earlier findings regarding accountability effects on multi-cue judgment accuracy. However, these tests are exploratory, as our hypotheses exclusively refer to the effectiveness of the experimental manipulation.

### **Dependent Variables and Manipulation Checks**

In order to test each of the three steps needed for an effective manipulation for both process and outcome accountability, we employed a series of instruction checks (step 1), as well as psychological manipulation checks (step 2) and psychological measures (step 3). We compared the results between the two accountability conditions and the no accountability condition.

*Instruction checks.* To check for the basic efficacy of our accountability manipulations, participants in all three conditions were asked on a 7-point Likert scale how likely they felt it would be that they would have to justify themselves (1 = not at all; 7 = very). Similar items have been used in the past by accountability researchers (e.g., Moser et al., 2020; Siegel-Jacobs & Yates, 1996; Tetlock et al., 1996). In addition, we used the instruction check reported by Häusser et al. (2017), in which participants in the two accountability conditions answered a single choice question regarding the content of the interview announced to them, indicating whether the interview would focus on the process, the results, or both equally. This single-choice question was repeated after participants had completed the experimental task to test the manipulation's persistence.

*Psychological manipulation checks.* To measure to what extent participants felt accountable for the process and the outcomes, respectively, we used a manipulation check previously employed by Schulz-Hardt et al. (2021). It asked participants to indicate whether they would focus more on the outcome of their choices or the process for making their choices during the task, and had to be answered on a 7-point Likert scale (1 = process; 7 = outcome). Participants answered this question a second time after they had completed the experimental task.

Additionally, we employed the manipulation check for process and outcome accountability that Zhang and Mittal (2005) developed to quantify our participants' felt process and outcome accountability more precisely. This measure consists of four items with a 7-point Likert scale for each process and outcome accountability. We adapted the items from Zhang and Mittal (2005) to our experimental task by slightly adjusting their wording and translating the items to German. Participants answered these items after completing the experimental task; in this, we follow the procedure of Zhang and Mittal (2005).

*Epistemic motivation.* Similar to previous accountability studies that measured epistemic motivation (e.g., De Dreu et al., 2006; de Langhe et al., 2011), we employed questionnaires to measure systematic information processing and rational thinking, as these two variables have been used as indicators of epistemic motivation in these previous studies. This practice is based on the idea that higher epistemic motivation should generally cause individuals to think more rationally and systematically about a given problem<sup>4</sup>. Following de Langhe et al. (2011), we slightly modified the situation-specific rational thinking style scale, validated by Novak and Hoffmann (2009), to assess epistemic motivation. This questionnaire consists of 10 statements that had to be answered on a 5-point Likert scale (1 = definitely

---

<sup>4</sup> This is of course a somewhat indirect measure of epistemic motivation and not necessarily ideal. Nonetheless, it is the way in which it has been assessed in past accountability research. Therefore, we decided to follow this practice as well instead of venturing into uncharted territory by employing a new measure of epistemic motivation

false, 5 = definitely true). We modified it by translating the items to German and slightly adjusted their wording to fit the specifics of our multi-cue estimation task. In addition to the situation-specific rational thinking style questionnaire, we employed another questionnaire that measured systematic information processing with three 5-point Likert scale items. This questionnaire also has been used previously to measure epistemic motivation, for example, by De Dreu et al. (1999) and De Dreu et al. (2006). Participants answered both these questionnaires after they had completed the experimental task; in this, we followed the procedures from De Langhe et al. (2011) and De Dreu et al. (1999).

*Stress.* To measure the perceived stress level of our participants, we followed the general procedure of Häusser et al. (2017), which, to our knowledge, is the only other study that ever empirically tested stress as a possible mediator for the effects of process and outcome accountability. However, we did deviate from Häusser et al. (2017), in that, instead of visual analog scales (VAS, e.g., Wewers & Lowe, 1990) to measure stress levels, we employed the state anxiety subscale from the German version of the STAI (Laux et al., 1981), because this scale could be better implemented in the online setting of this experiment. Not only has this scale been regularly used as a measure of psychological stress in the past (e.g., Glenk et al., 2020; Scholz et al., 2009), it also is highly correlated with VAS scores (e.g., Abend et al., 2014; Ducoulombier et al., 2020; Santamaría et al., 2022). The state anxiety subscale of the STAI consists of 20 items that are answered on 4-point Likert scales (1 = not at all; 4 = very) and was measured at two instances to measure changes in stress levels following the accountability manipulation. We measured state anxiety once at the beginning of the experiment, right after participants had signed the informed consent, and we measured it a second time after participants had finished the experimental task.

*Accuracy.* As an exploratory measure, we included the mean absolute percentage error (MAPE) participants made in their estimates of the package holiday prices. This measure

gives the accuracy of the price estimates in the form of the absolute deviation from the true prices in relation to the size of the true price. The more the estimates of a participant deviate from the correct prices, the larger the MAPE would become without an upper bound.

## Results

For both of our experiments, all data processing and all our analyses were conducted with the statistical package R (R core team, 2022). An overview of the results can be found in Table 1 and Table 2. All data and experimental materials can be retrieved from our OSF repository (<https://osf.io/f7wj9/>).

### Instruction checks

As mentioned above, the first step needed for a successful manipulation is that participants must understand and be aware of the manipulation. To test whether participants understood the instructions, we asked them how likely they felt it would be that they would have to justify themselves. We analyzed this question with a one-way ANOVA, which revealed a significant effect of condition,  $F(2, 233) = 10.30, p < .001, \eta^2 = .081$ . Post hoc tests showed that participants in the process accountability condition considered it to be significantly more likely that they would have to justify themselves ( $M = 4.94, SD = 1.60$ ) than participants in the no accountability ( $M = 3.76, SD = 1.75, t(233) = 4.08, p < .001, d = 0.65$ ) and participants in the outcome accountability condition ( $M = 3.74, SD = 2.06, t(233) = 3.91, p < .001, d = 0.66$ ). The latter two conditions did not significantly differ ( $t(233) = 0.10, p = .995, d = 0.01$ ).

As a second instruction check, we asked participants in the two accountability conditions what their interview would be about. A  $\chi^2$ - Test revealed that the distribution of the participants' answers significantly differed from a random distribution,  $\chi^2(2) = 48.83, p < .001$ . The majority of participants in the process accountability condition answered correctly that the interview would be about their decision-making process (47 of 67 = 70%), with only

a small percentage answering wrongly that the interview would be about their decision outcomes (2 of 67 = 3%), and the remaining participants answering that the interview would be about both about the outcome and the process (18 of 67 = 27%). In the outcome accountability condition, the majority of participants also answered correctly that the interview would be about their outcome (38 of 72 = 53%), in line with Hypothesis 1b. However, it should be noted that almost half of the participants did not answer entirely correct, and 15 out of 72 (21%) even answered utterly wrong that the interview would be about the process (the remaining 19 participants in the outcome accountability answered that the interview would be about both the process and the outcome and, therefore, were only partially incorrect in their answer).<sup>5</sup>

We asked participants again what they would be interviewed about after they finished the experimental task. Their answers were virtually identical to the first ones, with the distribution of the answers being significantly different from a random distribution,  $\chi^2(2) = 41.13, p < .001$ . Once again, the majority of the participants answered correctly, with 47 of 67 participants (70%) in the process accountability condition answering that the interview would be about their process, and 35 of 72 participants (49%) in the outcome accountability condition answering the interview would be about their outcomes.<sup>6</sup>

Therefore, we can conclude that participants in the process accountability condition understood and were aware of the manipulation. Accordingly, the manipulation did clear the first hurdle on the course to an effective manipulation of process accountability. In contrast, a

---

<sup>5</sup> We reran all of our analyses with only those participants who answered this question correctly. While results regarding the different instruction checks and the measures of felt accountability somewhat changed, in that the differences between the different conditions were more pronounced, there was no qualitative change in results for either measure of epistemic motivation or stress. All of these additional analyses can be found in Supplement B.

<sup>6</sup> In the process accountability condition, 3 participants (4%) answered completely wrong that the interview would be about their outcomes, and 17 participants (25%) answered it would be about both the outcome and the process equally. In the outcome accountability condition, 17 participants (24%) answered completely wrong, and 20 participants (28%) answered partially wrong that the interview would focus on outcome and process equally.

substantial amount of participants in the outcome accountability condition seemed to have problems understanding the manipulation and were somewhat less aware of it than the participants in the process accountability condition. Hence, we have to conclude that this manipulation was not particularly successful in taking the first step towards an effective manipulation of outcome accountability.

### **Psychological manipulation checks**

To check the second step of an effective manipulation, we employed a couple of measures to check whether or not our manipulation induced felt process and outcome accountability.

First, we asked participants what they would focus on during the task. A one-way ANOVA revealed a significant effect of condition,  $F(2, 233) = 6.81, p < .001, \eta^2 = .055$ . Post hoc tests revealed that outcome-accountable participants anticipated to focus more strongly on the outcome (compared to the process) ( $M = 4.44, SD = 1.57$ ) than process-accountable participants ( $M = 3.54, SD = 1.40$ ),  $t(233) = 3.65, p < .001, d = 0.62$ , and the participants in the no accountability condition ( $M = 3.91, SD = 1.42$ ),  $t(233) = 2.36, p = .019, d = 0.37$ . The latter two conditions did not significantly differ from each other,  $t(233) = 1.59, p = .113, d = 0.25$ .

We asked the same questions again after participants had completed the experimental task. The one-way ANOVA showed a significant effect of experimental condition,  $F(2, 233) = 9.08, p < .001, \eta^2 = .072$ . However, post hoc tests revealed slightly different results in the pairwise comparisons when compared to the first measurement. This time, the process-accountable participants answered that they had focused more on the process ( $M = 3.85, SD = 1.48$ ) than the participants in the other two conditions (outcome accountability:  $M = 4.83, SD = 1.51, t(233) = 3.85, p < .001, d = 0.65$ ; no accountability:  $M = 4.72, SD = 1.51, t(233) =$

3.65,  $p < .001$ ,  $d = 0.58$ ), while there was no difference between outcome-accountable and non-accountable participants,  $t(233) = 0.48$ ,  $p = .633$ ,  $d = 0.07$ .

To further test if our manipulation induced the psychological state of felt accountability, we employed the measures developed by Zhang and Mittal (2005).

*Felt outcome accountability.* We first analyzed to what extent the participants felt outcome accountability. For this measure, a one-way ANOVA revealed a significant effect of experimental condition,  $F(2, 233) = 3.66$ ,  $p = .027$ ,  $\eta^2 = .030$ . Participants in the outcome accountability condition reported feeling a stronger outcome accountability ( $M = 4.51$ ,  $SD = 0.97$ ) than participants in the process accountability condition ( $M = 4.11$ ,  $SD = 1.17$ ),  $t(233) = 2.61$ ,  $p = .010$ ,  $d = 0.44$ , and participants in the no accountability condition did ( $M = 4.24$ ,  $SD = 0.85$ ),  $t(233) = 1.98$ ,  $p = .049$ ,  $d = 0.31$ . There was no significant difference between the latter two conditions,  $t(233) = 0.86$ ,  $p = .393$ ,  $d = 0.14$ .

*Felt process accountability.* We conducted a similar one-way ANOVA for the participants' felt process accountability. This analysis found a significant effect of experimental condition,  $F(2, 233) = 13.21$ ,  $p < .001$ ,  $\eta^2 = .102$ . Post hoc tests revealed that process-accountable participants felt higher levels of process accountability ( $M = 4.52$ ,  $SD = 1.02$ ) than outcome-accountable participants ( $M = 3.83$ ,  $SD = 1.06$ ),  $t(233) = 4.13$ ,  $p < .001$ ,  $d = 0.70$ , and non-accountable participants ( $M = 3.76$ ,  $SD = 0.95$ ),  $t(233) = 4.83$ ,  $p < .001$ ,  $d = 0.77$ , whereas outcome-accountable and non-accountable participants did not differ significantly in their felt process accountability,  $t(233) = 0.42$ ,  $p = .672$ ,  $d = 0.07$ .

Taken together, these findings indicate that the manipulation also fulfilled the second requirement to be an effective manipulation of process accountability, as process-accountable participants did report higher levels of felt process accountability and answered that they focused more on their decision-making process while working on the experimental task than participants in the other two conditions. Even though we found somewhat equivocal evidence

regarding the question whether the manipulation fulfilled the first step towards a successful manipulation of outcome accountability, it did clear the second hurdle, as outcome-accountable participants felt more accountable for the outcome and answered that they planned to focus more on their outcomes during the experimental task.

### **Psychological consequences**

For the third and last step toward an effective manipulation, we analyzed our participants' epistemic motivation and stress levels in the three experimental conditions, as these are the two main mechanisms proposed in the literature on accountability as psychological consequences of process and outcome accountability.

*Epistemic motivation.* In order to test for the first psychological mechanisms proposed in the accountability literature, we used two instruments to measure epistemic motivation (namely the situation specific rational thinking style scale, Novak & Hoffmann, 2009, and the systematic information processing scale, DeDreu et al., 1999). Regarding the situation specific rational thinking style, a one-way ANOVA found no significant differences between experimental conditions,  $F(2, 233) = 0.60, p = .548, \eta^2 = .005$ . The same was true for the systematic information processing scale,  $F(2, 233) = 1.15, p = .319, \eta^2 = .010$ .

*Stress.* The second psychological mechanism which has been proposed to convey the effects of accountability is that different forms of accountability affect stress levels in different ways. To check for this possibility, we measured the participants' state anxiety with the corresponding subscale of the STAI, and we did so once before and once after the manipulation of accountability. As expected, we found no differences between experimental conditions in the baseline measure of state anxiety,  $F(2, 233) = 0.71, p = .494, \eta^2 = .006$ . When we measured the participants' state anxiety again after the task, we still did not find any significant differences between experimental conditions,  $F(2, 233) = 0.24, p = .792, \eta^2 = .002$ . In addition, a one-sample t-test against zero showed that there also were no significant



changes in state anxiety from the baseline to the second measure,  $t(235) = 0.71, p = .481, d = 0.05$ , and a repeated-measures ANOVA also found no significant main or interaction effects (all  $F$ s  $< 0.46$ , all  $p$ s  $> .541$ , all  $\eta^2 < .004$ ).

In conclusion, the interview-manipulation failed to fulfil the last requirement for an effective manipulation of process and outcome accountability, as in neither of the two conditions did it produce any of the psychological effects that the previous literature on accountability has proposed it should produce, i.e., increased in epistemic motivation and/or increased levels of stress.

### Exploratory analysis of accuracy

Finally, we ran an exploratory analysis on the effects of our manipulations on the mean absolute percentage error (MAPE) our participants made when estimating the prices of the package holidays – in other words, we tested for effects of the accountability manipulations on the participants' performance. While we had no hypotheses regarding the effects of process and outcome accountability on the accuracy of the price estimates participants made in the experimental task, based on the existing literature, it can be assumed that process accountability should lead to a decrease in MAPE-scores, as process accountability has been reported to improve judgment and decision quality. However, in a one-way ANOVA, we found no significant effect of experimental condition on MAPE-scores,  $F(2, 233) = 0.96, p = .385, \eta^2 = .008$ .

Table 1. Absolut frequencies of answers to the question regarding the focus of the interview announced to participants in the two accountability conditions (relative frequencies in parentheses).

Interview focus	Outcome accountability		Process accountability	
	Pre-task	Post-task	Pre-task	Post-task
Outcome	38 (53%)	35 (49%)	2 (3%)	3 (4%)
Process	15 (21%)	17 (24%)	47 (70%)	47 (70%)
Both equally	19 (26%)	20 (28%)	18 (27%)	17 (25%)

Table 2. Mean values of the dependent variables for the three experimental conditions in Experiment 1 (standard deviations in parentheses).

Variables	No accountability (n = 97)	Outcome accountability (n = 74)	Process accountability (n = 69)
<i>Instruction check</i>			
Likelihood of justification	3.76 (1.75) <sup>a</sup>	3.73 (2.06) <sup>a</sup>	4.94 (1.60) <sup>b</sup>
<i>Psychological manipulation checks</i>			
Planned task focus	3.91 (1.42) <sup>a</sup>	4.44 (1.57) <sup>b</sup>	3.54 (1.40) <sup>a</sup>
Reported task focus	4.72 (1.51) <sup>a</sup>	4.83 (1.51) <sup>a</sup>	3.85 (1.48) <sup>b</sup>
Felt outcome accountability	4.24 (0.85) <sup>a</sup>	4.54 (0.97) <sup>b</sup>	4.11 (1.17) <sup>a</sup>
Felt process accountability	3.76 (0.95) <sup>a</sup>	3.83 (1.06) <sup>a</sup>	4.53 (1.02) <sup>b</sup>
<i>Psychological mechanisms</i>			
SSTS	3.59 (0.68) <sup>a</sup>	3.47 (0.57) <sup>a</sup>	3.52 (0.67) <sup>a</sup>
Systematic information processing	3.54 (0.80) <sup>a</sup>	3.47 (0.82) <sup>a</sup>	3.67 (0.73) <sup>a</sup>
Baseline state anxiety	39.47 (9.54) <sup>a</sup>	40.26 (9.29) <sup>a</sup>	38.45 (7.83) <sup>a</sup>
Post-task state anxiety	39.11 (10.58) <sup>a</sup>	39.81 (10.05) <sup>a</sup>	38.69 (8.14) <sup>a</sup>
Δ state anxiety	-0.36 (4.77) <sup>a</sup>	-0.46 (4.62) <sup>a</sup>	0.24 (5.04) <sup>a</sup>
<i>Exploratory measures</i>			
MAPE	71.06 (67.91) <sup>a</sup>	69.06 (55.02) <sup>a</sup>	60.04 (45.80) <sup>a</sup>

Notes. Means not sharing a similar subscript differ per row at  $p < .05$  according to pairwise two-samples  $t$ -tests.

## Discussion

Since recent studies called into question the robustness of the effects of process accountability as a tool for debiasing and improving judgments and decision-making, our goal with this first experiment was to test whether the most commonly used manipulation of accountability works as intended. That is, we wanted to test whether the very common interview-manipulation of process and outcome accountability is understood by participants, whether it induces the psychological state of (different types of) felt accountability, and finally, whether this state produces the psychological effects proposed in the literature, namely increased epistemic motivation and increased stress levels.

Regarding the comprehension and salience of the manipulation (the instruction check), we found somewhat mixed results. The majority of participants in both the process and the outcome accountability condition correctly answered questions regarding what the interview that was announced to them would be about. These results statistically confirm Hypotheses 1a and 1b. However, it should be noted that the number of participants who misidentified what the interview announced to them would be about was relatively high, especially in the outcome accountability condition, where about half of the participants answered wrong. This descriptive difference in the number of correct answers between participants in the process and in the outcome accountability condition could, of course, simply be random fluctuation, as we have no reason to assume that the manipulation should be harder or easier to understand in one of the two accountability conditions. What is particularly astonishing, though, is that the high number of misidentifications of the interview content was already present in the first instruction check, that is, right after the manipulation of accountability.

Additionally, while process-accountable participants did feel it was more likely that they would have to justify themselves for their decision when compared to participants in the no accountability condition, no such difference was found between outcome-accountable and

non-accountable participants. Accordingly, we only found evidence supporting Hypothesis 2a, while Hypothesis 2b was not confirmed. Thus, while we found that our participants were mostly aware of and understood the interview-manipulation when used to induce process accountability, the outcome accountability manipulation was less successful.

For the second step in a successful experimental manipulation, namely that the manipulation inducing the psychological state that it aims at, we found that participants in the outcome accountability condition, even though they did not seem to be particularly aware of the manipulation, felt more accountable for the outcome than participants in the other two conditions. Similarly, participants in the process accountability condition did feel more accountable for the process than participants in the other two conditions. Consequently, we can confirm both Hypothesis 3 and Hypothesis 4 and, thus, conclude that the interview-manipulation of accountability does, indeed, induce the psychological state of feeling process or outcome-accountable. It should be noted, however, that the effect of the outcome accountability manipulation on the reported feelings of outcome accountability was relatively small ( $d = 0.31$ ), especially when we compare it to effects usually observed in psychological manipulation checks, and also compared to the effect of the process accountability manipulation on the reported feelings of process accountability ( $d = 0.77$ ). Despite this, the intended task focus was only affected for participants in the outcome accountability condition, who reported that they would focus more on the outcome when compared with the other two conditions. In contrast, process-accountable participants reported similar intentions with regard to their task focus as participants who were not held accountable at all.

Finally, when we checked the third step of an effective manipulation, we found that the interview-manipulation failed to take this step for both process and outcome accountability, as we did not find any indication that the manipulation was able to activate the proposed psychological mechanisms. Neither process nor outcome-accountable participants

showed any meaningful differences in epistemic motivation or stress level compared to non-accountable participants. Therefore Hypotheses 5 – 7 could not be confirmed.

From these results, we infer that the very commonly used interview-manipulation of accountability is generally salient to participants when used to induce process accountability – although the rate of participants to whom it is not salient is still too high, at least in our view – and does indeed produce feelings of being process-accountable. Despite the only partially successful instruction checks for outcome-accountable participants, we found some indications that the interview-manipulation was still successful in inducing some level of felt outcome accountability. However, contrary to the theoretical considerations and some earlier findings, the manipulation did not seem to have any effect on the psychological mechanisms (epistemic motivation and stress) that have been assumed to mediate the effect of process and outcome accountability on judgment and decision quality. These results are particularly worrisome because of the fairly high power of Experiment 1 (.94), which means that the likelihood that these findings are a type II error is relatively low<sup>7</sup>.

In sum, our findings question the notion that announcing an interview about participants' decision-making process or outcomes is a reliable way to produce the predicted effects of both accountability types on psychological mediators or judgment and decision quality. However, some factors might limit the generalizability of our findings in Experiment 1. The first limitation is that, in contrast to most other accountability studies, our first experiment was conducted as an online experiment, with an interview that was supposed to take place in the form of a video call. While we do not believe that this played a significant role in the results of our study, there is certainly a possibility that the online format could affect the results. For one, participants in an online experiment are not present in the

---

<sup>7</sup> This test power of .94 is achieved for a medium effect size, meaning it is of course lower for smaller effect sizes. However, given the fact that we are discussing psychological manipulation checks of the interview-manipulation any effect caused by the manipulation on the manipulation check that is not at least medium sized would mean that the manipulation would be rather unusable.

laboratory and do not have to make an effort to commute to the laboratory. The ease of participation, in turn, could lead to being less committed to participating, and to higher amounts of dropouts, especially in the two accountability conditions. This factor was very likely at play in our experiment, as only about 34% (241) of the 711 people who were interested in participating in the experiment (i.e., who opened the link to the experiment, read its landing page, which included information about its contents and procedure, and moved on beyond this first page) finished working on it. While this is a general phenomenon in online experiments, it is concerning to us that of the 470 people that began working on the experiment but later discontinued working on it, 82% (385) did so after they were assigned to one of the two accountability conditions<sup>8</sup>. Therefore, despite trying to limit selective dropouts by asking all participants to only participate if they were willing to be interviewed, we had substantial selective dropouts in this experiment. Such selective dropouts could, in turn, have skewed our results, as it is a possibility that especially those participants discontinue their participation who felt especially stressed by the accountability manipulation, which would then, in turn, hide the effect of the manipulation on participants' stress levels. It is, however, also possible that only those participants dropped out who were less committed to the experiment and, thus, would have been less affected by the accountability manipulation if they had continued. Therefore, it is also plausible that these dropouts could have even strengthened the effects of the accountability manipulation.

Another possible limitation is that most of the psychological effects of the accountability manipulation were measured *after* participants had finished working on the experimental task. While generally, we would expect any meaningful psychological states and effects that the accountability manipulation would induce to persist until after working on the

---

<sup>8</sup> Of the remaining 85 people that were interested in participating but did not finish the experiment, 76 quit before the manipulation took place, and 9 discontinued their participation after they were assigned to the no accountability condition.

experimental task, it is nevertheless possible that these effects change by working on the task itself. It has to be noted, however, that this is a rather typical limitation in accountability research, as psychological measures and manipulation checks, if they are measured at all, are usually measured after the participants completed working on the experimental task at hand (e.g., De Dreu et al., 1999; De Langhe et al., 2011; Zhang & Mittal, 2005). Hence, in applying these measures after working on the task was over, we followed the standards established in the literature so far.

Finally, a third limiting factor for the generalizability of our findings might be that, as mentioned above, we warned all participants at the start of the experiment that they might be interviewed. We made this decision in order to reduce selective dropouts in the accountability conditions. However, it is possible that this resulted in participants in the no accountability condition also feeling some form of accountability, as they might – at least in the beginning – have expected that the announcement of an interview might still come. This expectation could, for example, have increased epistemic motivation and stress levels among the non-accountable participants.

Because of these limitations, we decided to run a second experiment in the laboratory. We were aware of the fact that, due to the ongoing Covid-19-crisis and the measures that have been taken in order to counteract it (e.g., online teaching in many university courses), such a laboratory experiment would necessarily have to have a much smaller sample than the online experiment. Nevertheless, it should still allow us to check whether the pattern of results substantially differs between the two study modes.

## **Experiment 2**

We ran Experiment 2 to address the possible limitations to the generalizability of our findings in Experiment 1. To this end, we changed the setting of the experiment. While Experiment 1 was conducted as an online study, Experiment 2 was conducted in the

laboratory. This is, one, more typical for research on accountability and, two, should solve the problem that we had with selective dropouts in Experiment 1, as from our own experience (cf. Schulz-Hardt et al., 2021; Rollwage et al., 2022), (selective) dropouts are not very common in laboratory studies on accountability. Furthermore, we changed the order of the measures we employed in Experiment 2 as, this time, the psychological manipulation checks were conducted before participants worked on the experimental task. Finally, in this experiment, we explicitly told participants in the no accountability condition that they would not be interviewed about their decisions. Apart from these changes, Experiment 2 was identical to Experiment 1.

## **Method**

### **Participants and Design**

A total of 68 participants (44 female, 24 male), ranging in age from 19 years to 60 years ( $M = 26.9$ ,  $SD = 8.04$ ), were recruited using the Online Recruitment System for Economic Experiments (ORSEE; Greiner, 2015) and participated in Experiment 2 at our laboratory at the University of Goettingen. Participants in Experiment 2 were randomly assigned to one of three experimental conditions in a one-factorial between-subject design (no accountability vs. process accountability vs. outcome accountability). We compensated participants with either a 5€ show-up fee or by earning partial course credit; furthermore, all participants had the chance to enter a cash raffle in which they could win 10 x 20€ and 10 x 10€. Given that Experiment 2 also was conducted during the Covid-19 pandemic, we had to severely limit the amount of people that we simultaneously test in our laboratory, which is why we only were able to collect a relatively small sample in this second experiment.

### **Procedure and Materials**

Participants were invited to the laboratory in groups of up to 8 persons. Upon entering the lab, they were greeted by a research assistant and seated in front of a laptop. The



experiment, which was constructed with the alfred3 Software (Treffenstädt et al., 2021), was already opened on the laptop screen. With a few exceptions, described below, the experiment was identical to Experiment 1. Participants were informed to follow the instructions on the laptop screen and that, if anything was unclear to them, they should raise their hand so that the research assistant could answer their questions. First, participants read the informed consent form to which they had to agree to further participate in the experiment. After they gave their consent, participants filled in the state anxiety subscale of the STAI (Laux et al., 1981) to measure their baseline stress level. Next came the manipulation of accountability and most of the measures of interest (these included instruction checks, measures of felt accountability, and the two measures of epistemic motivation). Then participants began to work on a task estimating package holiday prices (the same task as in Experiment 1). After participants finished the experimental task, they filled in a final questionnaire, consisting of the second measure of stress, a repeat of the instruction checks, a suspicion check, and some demographics, which concluded the experiment. Finally, participants were thanked for their participation and fully debriefed by the research assistant.

### **Manipulation of accountability**

Again, before the start of the experiment, all participants were informed that some of them would be randomly chosen to participate in an interview with a researcher and an experienced travel clerk. At a later point in the experiment, participants in the two accountability conditions saw a screen that announced to them that they had been chosen for an interview and were requested to enter two dates and times in the next two weeks at which they would be available for the interview, together with their e-mail address in order to schedule the interview. In contrast to this, participants in the no accountability condition saw a screen that explicitly informed them that they had not been chosen to be interviewed.

### **Multi-cue estimation task**

We used the same multi-cue estimation task of package holidays as in Experiment 1 for the experimental task.

### **Dependent Variables and Manipulation Checks**

We employed the same measures and dependent variables as in Experiment 1. The only difference was that, this time, some of the measures were taken *before* participants worked on the multi-cue estimation task (instead of being taken afterward, as in Experiment 1). Therefore, we had to slightly change some of the wording, so that the items were formulated prospectively instead of retrospectively. The instruments that were used before the experimental task were the measures of felt process and outcome accountability (Zhang & Mittal, 2005) as well as the two scales for epistemic motivation (SSTS, Novak & Hoffmann, 2009; systematic information processing, DeDreu, 1999).

### **Results**

An overview of descriptive statistics and our results for Experiment 2 can be found in Table 3 and Table 4.

### **Instruction checks**

We analyzed our instruction checks as the first step in testing the effectiveness of the interview-manipulation of accountability. First, we ran a one-way ANOVA on the question of how likely participants felt it would be that they would have to justify themselves. We found a significant effect of experimental condition,  $F(2, 65) = 10.30, p = .007, \eta^2 = .14$ . Post hoc tests revealed that participants in the process accountability condition considered it to be significantly more likely that they would have to justify themselves ( $M = 4.86, SD = 1.55$ ) than participants in the no accountability ( $M = 3.24, SD = 1.76$ ),  $t(65) = 3.16, p = .002, d = 0.96$ , and participants in the outcome accountability condition ( $M = 3.72, SD = 1.74$ ),  $t(65) =$

2.32,  $p = .024$ ,  $d = 0.68$ , while the latter two conditions did not differ significantly from one another,  $t(65) = 0.96$ ,  $p = .339$ ,  $d = 0.29$ .

As a second instruction check, we asked participants in the two accountability conditions what the focus of the announced interview would be. A  $\chi^2$ - Test revealed that participants answers were not randomly distributed,  $\chi^2(2) = 6.82$ ,  $p = .043$ . A substantial number of participants in both the outcome accountability condition and the process accountability condition answered that the interview would focus both on their outcomes and their process equally (PA: 11 of 22 = 50%; OA: 11 of 25 = 44%), while only eight of the participants in the outcome accountability condition (32%) and ten participants in the process accountability condition (45%) answered correctly that the interview would be exclusively about the outcome, or exclusively about the process, respectively.

We repeated this second instruction check after participants had completed working on the experimental task. The answers participants gave here were similar to their initial answers. The distribution of the answers was again significantly different from a random distribution,  $\chi^2(2) = 6.98$ ,  $p = .031$ . Eleven of the 22 participants (50%) in the process accountability condition answered correctly that the interview would focus on the process. However, a substantial part of participants in this condition (10 of 22 = 45%) still thought the interview would focus on outcomes and processes equally. In contrast, in the outcome accountability condition, 11 of the 25 participants (44%) answered that the interview would focus on their outcomes and their process equally, while only seven participants answered correctly that the interview would be about their outcomes (28%).

From this, we can conclude that participants in the process accountability condition understood the manipulation reasonably well and were somewhat aware of it, while, participants in the outcome accountability condition neither fully understood the manipulation

nor were they particularly aware of it. Therefore, our manipulation did not fully clear the first hurdle toward an effective manipulation of accountability.

### **Psychological manipulation checks**

Next, we checked the second step of an effective manipulation. That is, we checked whether or not the manipulation of process and outcome accountability induced different levels of felt process and outcome accountability, as well as differences in what participants focused on during the task.

First, we ran a one-way ANOVA on the question regarding participants' planned focus during the task. This analysis showed no significant effect of experimental condition,  $F(2, 65) = 0.07, p = .929, \eta^2 = .002$ .

However, another one-way ANOVA on the question regarding participants' retrospectively reported focus during the task showed a significant effect of experimental condition,  $F(2,65) = 3.34, p = .042, \eta^2 = .093$ .

Post hoc tests revealed that outcome-accountable participants answered that they had focused more on the outcome ( $M = 4.88, SD = 1.45$ ) than the participants in the other two conditions (PA:  $M = 3.91, SD = 1.87, t(65) = 2.17, p = .034, d = 0.63$ ; no accountability:  $M = 3.86, SD = 1.20, t(65) = 2.25, p = .028, d = 0.67$ ), while there was no significant difference between process-accountable and non-accountable participants,  $t(65) = 0.11, p = .912, d = 0.03$ .

*Felt outcome accountability.* Whether our manipulations induced the psychological state of felt accountability was measured with the instrument developed by Zhang and Mittal (2005). For the measure of felt outcome accountability, we calculated a one-way ANOVA, which found no significant effect for condition,  $F(2, 65) = 0.03, p = .967, \eta^2 = .001$ .

*Felt process accountability.* For the measure of felt process accountability, we calculated another one-way ANOVA, which found that the experimental conditions significantly affected felt process accountability,  $F(2, 65) = 4.25, p = .018, \eta^2 = .116$ . Afterward, we calculated post hoc tests to test for differences between the individual conditions. These revealed that process-accountable participants felt higher levels of process accountability ( $M = 4.52, SD = 0.98$ ) than outcome-accountable participants ( $M = 3.79, SD = 0.73$ ),  $t(65) = 2.84, p = .006, d = 0.83$ . Participants in the no accountability condition ( $M = 4.29, SD = 0.94$ ) did not significantly differ from participants in the process accountability condition,  $t(65) = 0.88, p = .382, d = 0.27$ , or participants in the outcome accountability condition,  $t(65) = 1.90, p = .062, d = 0.56$ .

Based on these analyses, we conclude that the interview-manipulation was partially successful in inducing the psychological state of felt process accountability. In contrast, the manipulation did not affect felt outcome accountability. However, participants in the outcome accountability condition reported focusing more on the outcome during the experimental task than participants in the other two conditions. In summary, the interview-manipulation did not manage to clearly fulfil the second criterion for an effective manipulation of either process or outcome accountability.

### **Psychological consequences**

The third check for an effective manipulation consists of testing for the psychological consequences of the manipulation. For this step, we analyzed epistemic motivation and stress levels in our experimental conditions, as these are the two main mechanisms for accountability effects proposed in the literature.

*Epistemic motivation.* The two instruments we used to measure epistemic motivation were analyzed using two one-way ANOVAs. Here, we did not find any significant effect of experimental condition on epistemic motivation, regardless of whether we measured

epistemic motivation with the SSTS (Novak & Hoffmann, 2009),  $F(2, 65) = 0.34, p = .710, \eta^2 = .010$ , or the systematic information processing scale we adapted from DeDreu et al. (1999),  $F(2, 65) = 1.49, p = .232, \eta^2 = .044$ .

*Stress.* Regarding our measurements of perceived stress, we found no significant effects of experimental condition at the baseline measure of state anxiety,  $F(2,65) = 0.91, p = .400, \eta^2 = .028$ , or – more importantly – in the second measurement, which was taken directly after the experimental task,  $F(2, 65) = 1.49, p = .234, \eta^2 = .044$ . On average, there was a significant change in stress levels, as a one-sample  $t$ -test against zero showed that participants reported significantly lower state anxiety on the critical measure after the task,  $t(67) = 3.24, p = .002, d = .39$ , than on the baseline measure. However, there were no significant differences in this change between the experimental conditions,  $F(2, 65) = 0.30, p = .739, \eta^2 = .009$ . Furthermore, this change in state anxiety was, at least for the two accountability conditions, in the opposite direction of our predictions based on the previous literature, as the literature would predict an increase in anxiety as a consequence of the accountability manipulations.

To summarize, the interview-manipulation of process accountability and outcome accountability did not clear the third hurdle for any effective manipulation, as it affected neither epistemic motivation nor stress levels.

### **Exploratory analysis of accuracy**

Like in Experiment 1, we included the participants' accuracy as an additional exploratory analysis by calculating a one-way ANOVA on their MAPEs. This analysis found no significant differences between the experimental conditions,  $F(2, 65) = 0.58, p = .563, \eta^2 = .018$ .

Table 3. Absolut frequencies of answers to the question regarding the focus of the interview announced to participants in the two accountability conditions (relative frequencies in parentheses).

Interview focus	Outcome accountability		Process accountability	
	Pre-task	Post-task	Pre-task	Post-task
Outcome	8 (32%)	7 (28%)	1 (5%)	1 (5%)
Process	6 (24%)	7 (28%)	10 (45%)	11 (50%)
Both equally	11 (44%)	11 (44%)	11 (50%)	10 (45%)

Table 4. Mean values of the dependent variables for the three experimental conditions in Experiment 2 (standard deviations in parentheses).

Variables	No accountability (n = 21)	Outcome accountability (n = 25)	Process accountability (n = 22)
<i>Instruction check</i>			
Likelihood of justification	3.24 (1.76) <sup>a</sup>	3.72 (1.74) <sup>a</sup>	4.86 (1.55) <sup>b</sup>
<i>Psychological manipulation checks</i>			
Planned task focus	4.00 (1.22) <sup>a</sup>	4.16 (1.37) <sup>a</sup>	4.09 (1.60) <sup>a</sup>
Actual task focus	3.86 (1.20) <sup>a</sup>	4.88 (1.45) <sup>b</sup>	3.91 (1.87) <sup>a</sup>
Felt outcome accountability	4.31 (0.95) <sup>a</sup>	4.31 (0.83) <sup>a</sup>	4.38 (1.11) <sup>a</sup>
Felt process accountability	4.29 (0.94) <sup>ab</sup>	3.79 (0.73) <sup>a</sup>	4.52 (0.98) <sup>b</sup>
<i>Psychological mechanisms</i>			
SSTS	3.73 (0.75) <sup>a</sup>	3.60 (0.58) <sup>a</sup>	3.60 (0.52) <sup>a</sup>
Systematic information processing	4.14 (0.60) <sup>a</sup>	3.96 (0.72) <sup>a</sup>	3.80 (0.60) <sup>a</sup>
Baseline state anxiety	37.76 (6.80) <sup>a</sup>	36.00 (6.29) <sup>a</sup>	38.59 (7.00) <sup>a</sup>
Post-task state anxiety	35.62 (5.50) <sup>a</sup>	34.32 (6.76) <sup>a</sup>	37.45 (6.25) <sup>a</sup>
$\Delta$ state anxiety	-2.14 (4.14) <sup>a</sup>	-1.68 (4.34) <sup>a</sup>	-1.14 (4.21) <sup>a</sup>
<i>Exploratory measure</i>			
MAPE	80.94 (71.89) <sup>a</sup>	61.62 (56.75) <sup>a</sup>	66.17(59.07) <sup>a</sup>

Notes. Means not sharing a similar subscript differ per row at  $p < .05$  according to pairwise two-samples  $t$ -tests.



## Discussion

Experiment 2 was conducted to address some of the limitations of Experiment 1. We found that the results of Experiment 2 largely mirror those of Experiment 1. In Experiment 2, the instruction checks were, again, only partially successful, as the majority of participants could not correctly identify the supposed topic of their interview, and only process-accountable participants felt it would be more likely that they would have to justify themselves (compared to the control condition). In contrast, outcome-accountable participants did not feel it would be more likely that they would have to justify themselves compared to non-accountable participants. Furthermore, we found no difference regarding what participants planned to focus on while working on the experimental task. Thus, the path for the interview-manipulation of (process) accountability to have favorable effects on judgments and decision-making already goes partially astray on the very first step along the way, as the content of the manipulation does not seem sufficiently salient to participants.

Regarding the second step towards a successful manipulation, which is whether the manipulation induces the psychological state of felt accountability, we found that in Experiment 2, the manipulation was, at best, only partially successful, as it did not affect the level of felt outcome accountability. While process-accountable participants did feel more accountable for the process than outcome-accountable participants, they did not report higher felt process accountability than non-accountable participants.

The third and last step to a successful and robust manipulation is its ability to activate the psychological mechanisms behind the changes in the dependent variables. In the case of process accountability, these supposedly are epistemic motivation and stress. However, just like in Experiment 1, we found no indication that there were any effects of the manipulation on either epistemic motivation or stress in Experiment 2.

Experiment 2 offers further support to the conclusion that the widely used interview-manipulations of process and outcome accountability have rather diffuse or volatile effects, as it seems that participants often are not particularly aware of them. In addition, they seem to be unable to reliably activate the psychological processes that the literature claims they should activate.

Of course, one might argue that Experiment 2 did not find these effects because of its smaller sample size and, thus, its limited test power. However, even though the sample size is much smaller than the sample in Experiment 1, it is still sufficient to allow for reliable detection of large effects (power = .83 for  $f = 0.40$ ). If we assume that any reliable manipulation should have reasonably large effects on instruction and manipulation check items, as well as on measures of the psychological mechanism mediating its effects on other dependent variables, we should have found most of the relevant effects even with the reduced sample size of Experiment 2. Furthermore, if we focus on the estimated effect sizes in Experiment 2, the results are very similar to Experiment 1. Hence, effects that failed to become significant due to insufficient test power in Experiment 2 should have been significant in Experiment 1. The fact that this was hardly ever the case adds to our trust in the current findings. After all, there is little reason to believe that the procedural and contextual changes introduced in Experiment 2 lead to a more optimistic picture regarding the effects of the interview-manipulation of accountability.

## **General Discussion**

The present study tested how reliable and effective one of the most commonly used manipulation of accountability, namely announcing an interview to participants in which they have to either justify their decision-making process or their decision outcomes, is in creating the psychological effects it is supposed to produce. Specifically, we tested how well aware participants are of the content of the manipulation, whether this manipulation induces the

psychological states of felt (process and outcome) accountability, and whether the mediating psychological mechanisms proposed in the literature are affected.

Our two experiments showed essentially identical results regarding the effectiveness of the interview-manipulation of process and outcome accountability. Both Experiment 1 and Experiment 2 found evidence that participants in the process accountability conditions generally understood the manipulation and that the manipulation successfully induced felt process accountability. It has to be mentioned, however, that the number of process-accountable participants who were unable to correctly identify what they would be interviewed about was still surprisingly high (about 30% in Experiment 1 and around 50% in Experiment 2). For outcome accountability, the results were even more problematic, as only around 50% of outcome accountable participants were able to correctly identify the content of the interview in Experiment 1, and in Experiment 2 it was only a minority of participants in the outcome accountability condition who properly understood the manipulation. What makes these results even more surprising is that the first of these checks was done right after participants had read the manipulation, so this information should have been easy to recognize. This indicates to us that the manipulation was, in general, not particularly well understood by our participants, and that it was not very salient to them.

Regarding the effectiveness of the manipulation to induce felt outcome accountability, the results in our study were mixed. In Experiment 1, we found that the manipulation did indeed induce felt outcome accountability, whereas, in Experiment 2, we found no significant differences in the felt outcome accountability of our participants. It has to be said, however, that this difference might very well be caused by the lower power of Experiment 2, as the effect sizes for the differences in felt accountability between the experimental conditions were very similar in both experiments. Finally, both of our experiments also found identical results concerning the psychological mechanisms proposed in the literature, as neither experiment

found any significant effects of the accountability manipulation on either the measures of epistemic motivation or the measure of stress. In conclusion, while the interview-manipulation that is very prevalent in accountability research was, at least partially, able to clear the first two hurdles that have to be taken in order for a manipulation to be effective, it failed to clear the third and last hurdle, that is, it did not activate any of the proposed psychological mechanisms that it should have activated.

These results have important implications for our understanding of accountability effects and the research methods we use to investigate these effects. For one, our findings further question the reliability of the beneficial effects of process accountability that were reported in the literature but more recently have been called into question by a couple of unsuccessful replication attempts (e.g., Schulz-Hardt et al., 2021; Rollwage et al., 2022). The reason for this is that substantial parts of the literature on accountability rely on exactly the same manipulation of process and outcome accountability that we employed in the two experiments reported here, namely announcing an interview to participants in which they will have to justify their decision-making process or their decision outcomes, respectively (e.g., Girandola & Gauthier, 2001; Häusser et al. 2017; Siegel-Jacobs & Yates, 1996). As our findings indicate, it seems unlikely that this manipulation can reliably induce a form of process accountability that can produce the previously claimed effects on judgment and decision quality.

To be fair though, one of the psychological mechanisms that we tested, namely epistemic motivation, has been shown to be affected by process accountability manipulations several times in the past (e.g., De Dreu et al., 2006; Scholten et al., 2007; De Langhe et al. 2011). Thus, it would be premature to conclude that there is no connection between process accountability and epistemic motivation at all. However, what our findings do indicate is that

this connection, if it exists, is either much weaker than expected<sup>9</sup> or way more fragile, i.e., the effect may be highly dependent on contextual factors. Since, as of now, none of the hidden moderators that could potentially determine the effectiveness of the manipulation have been identified, using this manipulation in future studies should, at least for now, be avoided. As currently, one cannot be sure whether or not they achieved the needed configuration of hidden moderators for the manipulation to be effective.

## **Limitations and future research**

Our study contributes to the accountability literature by thoroughly checking for the effects of the most commonly used manipulation of process accountability. This is important because previous research often neglected manipulation checks, and while some psychological effects of the accountability manipulations have been proposed in the past, they have seldomly been empirically tested.

Nonetheless, our study has some limitations that should be addressed in future research. First, while we tried to match the typical manipulation as closely as possible, there are some factors that we had to slightly adapt for the use in the online setting of Experiment 1 and that we, for the sake of comparability, did not change in Experiment 2. For example, we announced an interview that would be conducted via an online video call, and this call would take place sometime in the next 14 days. This might have somewhat reduced the strength of our manipulation when compared with interviews that supposedly takes place right after the experiment and in the physical presence of the interviewers. It should be noted, however, that such a time lag between experimental task and interview is nothing particularly new or

---

<sup>9</sup> The effect sizes of the effect of accountability on epistemic motivation in our two experiments ranged from  $\eta^2 = .005$  to  $\eta^2 = .044$ , whereas the effect sizes in the previous literature range from about  $\eta^2 = .06$  to about  $\eta^2 = .11$ .

untypical in accountability research, as other studies have used similar procedures in the past, e.g., Simonson & Staw (1992), Scholten et al. (2007).

Another limitation is that, as in most previous studies investigating the effects of accountability, we only used self-report measures of stress and epistemic motivation. Such self-report measures can be, of course, prone to self-report biases (cf. Diener, 1984). While we do not expect that these biases, if they were at work in our experiments, would be correlated in any way with the different experimental conditions, they could still produce noise in our data and, thus, potentially affect the results of our analysis. In addition, it must be noted that stress is a more complex phenomenon than just an increase in anxiety, as research has shown that stress induces various physiological, psychological as well as emotional, and social effects (Frisch et al., 2015). Therefore, it is possible that just looking at changes in state anxiety might be too narrow in scope, and that we might have seen changes in stress based on our accountability manipulations had we broadened our scope to include more diverse stress indicators. Following from this, it would be a sensible way forward to employ more objective measures of stress, for example, by measuring physiological reactions such as cardiovascular responses (Blascovich & Mendes, 2000) or salivary cortisol levels (Hellhammer et al., 2009).

One final limitation that we want to mention is the fact that, in our experiments, we focused solely on those two psychological mechanisms that, so far, have been proposed in the accountability literature. While, of course, these two mechanisms should be the first ones to be tested, there is a chance that they are not the actual or at least not the only relevant mechanisms behind the effects of process and outcome accountability on judgments and decision-making. Therefore, it might be that process and outcome accountability both affect judgment and decision-making in the way the literature claims, but that these effects work through different, and as of now unidentified, pathways. We did not test for any such

alternative psychological mechanisms that could mediate the effect of process and outcome accountability on judgment and decision-making.

Despite these limitations, we are convinced that our results warrant the conclusion that researchers should approach investigating process and outcome accountability through new avenues, as the interview-manipulation of accountability that is usually used in lab experimental research seems to have, at best, weak and volatile effects. This conclusion is not only supported by the results of our two experiments reported above, but also by a number of recent replication attempts of studies that also used the interview-manipulation of accountability (e.g., Häusser & Schulz-Hardt, 2017; Schulz-Hardt et al., 2021; Rollwage et al. 2022). Therefore, one of the most important, if not *the* most important, implications for future research on process and outcome accountability is the necessity to find a more adequate manipulation of accountability. Since, as discussed earlier in this paper, the interview-manipulation of accountability is far removed from accountability in the wild, it would be crucial that any potential new paradigm aiming to manipulate process and outcome accountability more closely resembles real-world accountability. If such a realistic manipulation should be too hard to realize, another alternative avenue for future research on process and outcome accountability would be to concentrate on field studies in which the different kinds of accountability occur naturally and can be compared with each other and with situations without accountability.

## **Conclusion**

Our findings show that the most commonly used experimental manipulation of process and outcome accountability is not particularly salient to a substantial part of study participants and, more importantly, does not reliably trigger the psychological mechanisms that have been proposed to mediate accountability effects on judgment and decision quality. This further substantiates the evidence from previous studies questioning the robustness of reported effects

of accountability, and it implies that alternative research paradigms are needed if we want to identify effects of process and/or outcome accountability that are both robust and generalizable beyond the laboratory.



## References

- Abend, R., Dan, O., Maoz, K., Raz, S., & Bar-Haim, Y. (2014). Reliability, validity and sensitivity of a computerized visual analog scale measuring state anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, *45*(4), 447–453.  
<https://doi.org/10.1016/j.jbtep.2014.06.004>
- 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](https://doi.org/10.1016/0749-5978(85)90049-4)
- Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational behavior and human decision processes*, *37*(1), 93–110. [https://doi.org/10.1016/0749-5978\(86\)90046-4](https://doi.org/10.1016/0749-5978(86)90046-4)
- Blascovich, J., & Mendes, W. B. (2000). Challenge and threat appraisals: The role of affective cues. In J. Forgas (Hrsg.), *Feeling and Thinking: The Role of Affect in Social Cognition* (S. 59–82). Cambridge University Press.
- Bovens, M. A. P., Goodin, R., & Schillemans, T. (2014). *The Oxford Handbook of Public Accountability*. Oxford University Press.
- Brtek, M. D., & Motowidlo, S. J. (2002). Effects of procedure and outcome accountability on interview validity. *The Journal of Applied Psychology*, *87*(1), 185–191. <https://doi.org/10.1037/0021-9010.87.1.185>
- Chang, W., Atanasov, P., Patil, S., Mellers, B. A., & Tetlock, P. E. (2017). Accountability and adaptive performance under uncertainty: A long-term view. *Judgment and Decision Making*, *12*(6), 610–626.
- De Dreu, C. K. W., Beersma, B., Stroebe, K., & Euwema, M. C. (2006). Motivated information processing, strategic choice, and the quality of negotiated agreement.

*Journal of Personality and Social Psychology*, 90(6), 927–943.

<https://doi.org/10.1037/0022-3514.90.6.927>

De Dreu, C. K. W., Koole, S. L., & Oldersma, F. L. (1999). On the seizing and freezing of negotiator inferences: Need for cognitive closure moderates the use of heuristics in negotiation. *Personality & Social Psychology Bulletin*, 25(3), 348–362.

<https://doi.org/10.1177/0146167299025003007>

De Dreu, C. K., Koole, S. L., & Steinel, W. (2000). Unfixing the fixed pie: a motivated information-processing approach to integrative negotiation. *Journal of Personality and Social Psychology*, 79(6), 975–987. <https://doi.org/10.1037/0022-3514.79.6.975>

De Dreu, C. K. W., & van Knippenberg, D. (2005). The possessive self as a barrier to conflict resolution: effects of mere ownership, process accountability, and self-concept clarity on competitive cognitions and behavior. *Journal of Personality and Social Psychology*, 89(3), 345–357. <https://doi.org/10.1037/0022-3514.89.3.345>

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95(3), 542–575.

<https://doi.org/10.1037/0033-2909.95.3.542>

Ducoulombier, V., Chiquet, R., Graf, S., Leroy, B., Bouquet, G., Verdun, S., Martellier, F., Versavel, A., Kone, A., Lacroix, K., Duthoit, D., Lenglet, Q., Devaux, A., Jeanson, R., Lefebvre, A., Coviaux, B., Calais, G., Grimbert, A., Ledein, M., [...] Houvenagel, E. (2020). Usefulness of a Visual Analog Scale for measuring anxiety in hospitalized patients experiencing pain: A multicenter cross-sectional study. *Pain Management Nursing: Official Journal of the American Society of Pain Management Nurses*, 21(6), 572–578. <https://doi.org/10.1016/j.pmn.2020.03.004>

Ederer, F., & Manso, G. (2013). Is pay for performance detrimental to innovation?

*Management Science*, 59(7), 1496–1513, <https://doi.org/10.1287/mnsc.1120.1683>

- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Frisch, J. U., Häusser, J. A., & Mojzisch, A. (2015). The Trier Social Stress Test as a paradigm to study how people respond to threat in social interactions. *Frontiers in Psychology*, 6, 14. <https://doi.org/10.3389/fpsyg.2015.00014>
- Girandola, F., & Gauthier, E. (2001). Organizational decision making: Effects of accountability on escalation of commitment. *European Review of Applied Psychology = Revue Europeenne de Psychologie Appliquee*, 51(1), 111–119.
- Glenk, L. M., Kothgassner, O. D., Felnhofer, A., Gotovina, J., Pranger, C. L., Jensen, A. N., Mothes-Luksch, N., Goreis, A., Palme, R., & Jensen-Jarolim, E. (2020). Salivary cortisol responses to acute stress vary between allergic and healthy individuals: the role of plasma oxytocin, emotion regulation strategies, reported stress and anxiety. *Stress*, 23(3), 275–283. <https://doi.org/10.1080/10253890.2019.1675629>
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1), 114–125. <https://doi.org/10.1007/s40881-015-0004-4>
- Häusser, J. A., Frisch, J. U., Wanzel, S., & Schulz-Hardt, S. (2017). Effects of process and outcome accountability on idea generation. *Experimental Psychology*, 64(4), 262–272. <https://doi.org/10.1027/1618-3169/a000368>
- Häusser, J. A., & Schulz-Hardt, S. (2017). *Ergebnisbericht zum DFG-Projekt „Prozess- und Ergebnisverantwortlichkeit bei Gruppenentscheidungen“*. Universität Gießen & Universität Göttingen.

- Hellhammer, D. H., Wüst, S., & Kudielka, B. M. (2009). Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology*, *34*(2), 163–171.  
<https://doi.org/10.1016/j.psyneuen.2008.10.026>
- Hoffmann, J. A., Gaissmaier, W., & von Helversen, B. (2017). Justifying the judgment process affects neither judgment accuracy, nor strategy use. *Judgment and Decision Making*, *12*(6), 627–641.
- Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *The Journal of Consumer Research*, *9*(1), 90–98. <https://doi.org/10.1086/208899>
- Janis, I. L., & Mann, L. (1977). *Decision making: A Psychological Analysis of Conflict, Choice, and Commitment* (Bd. 488). Free Press.
- Keinan, G., Friedland, N., & Ben-Porath, Y. (1987). Decision making under stress: Scanning of alternatives under physical threat. *Acta Psychologica*, *64*(3), 219–228.  
[https://doi.org/10.1016/0001-6918\(87\)90008-4](https://doi.org/10.1016/0001-6918(87)90008-4)
- Kim, S., & Trotman, K. T. (2015). The comparative effect of process and outcome accountability in enhancing professional scepticism. *Accounting and Finance*, *55*(4), 1015–1040. <https://doi.org/10.1111/acfi.12084>
- Kruglanski. (1989). The psychology of being "right": The problem of accuracy in social perception and cognition. *Psychological Bulletin*, *106*(3), 395–409. <https://doi.org/10.1037/0033-2909.106.3.395>
- Langhe, B. de, van Osselaer, S. M. J., & Wierenga, B. (2011). The effects of process and outcome accountability on judgment process and performance. *Organizational Behavior and Human Decision Processes*, *115*(2), 238–252.  
<https://doi.org/10.1016/j.obhdp.2011.02.003>

- Laux, L., Glanzmann, P., Schaffner, P., & Spielberger, C.D. (1981). *Das State-Trait-Angstinventar. Theoretische Grundlagen und Handanweisungen*. Beltz.
- Lavedán Santamaría, A., Masot, O., Canet Velez, O., Botigué, T., Cemeli Sánchez, T., & Roca, J. (2022). Diagnostic Concordance between the Visual Analogue Anxiety Scale (VAS-A) and the State-Trait Anxiety Inventory (STAI) in Nursing Students during the COVID-19 Pandemic. *International Journal of Environmental Research and Public Health*, 19(12). <https://doi.org/10.3390/ijerph19127053>
- Lefebvre, M., & Vieider, F. M. (2013). Reining in excessive risk-taking by executives: the effect of accountability. *Theory and Decision*, 75(4), 497–517.  
<https://doi.org/10.1007/s11238-012-9335-2>
- Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255–275. <https://doi.org/10.1037/0033-2909.125.2.255>
- Novak, T. P., & Hoffman, D. L. (2009). The fit of thinking style and situation: New measures of situation-specific experiential and rational cognition. *The Journal of Consumer Research*, 36(1), 56–72. <https://doi.org/10.1086/596026>
- Patil, S. V., Vieider, F., & Tetlock, P. E. (2014). Process versus outcome accountability. *The Oxford Handbook of Public Accountability*, 69–89.  
doi:10.1093/oxfordhb/9780199641253.013.0002
- R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>. [computer software]
- Rollwage, J., Häusser, J. A., & Schulz-Hardt, S. (2022). *Does holding people accountable make them less loss averse? A three-study replication attempt*. [Manuscript submitted for publication], Department of Economic and Social Psychology, University of Goettingen.

- Schneider, S. L., & Shanteau, J. (Hrsg.). (2003). Emerging perspectives on judgment and decision research. *Cambridge Series on Judgment and Decision Making.*, 713.  
<https://doi.org/10.1017/CBO9780511609978>
- Scholten, L., van Knippenberg, D., Nijstad, B. A., & De Dreu, C. K. W. (2007). Motivated information processing and group decision-making: Effects of process accountability on information processing and decision quality. *Journal of Experimental Social Psychology*, 43(4), 539–552. <https://doi.org/10.1016/j.jesp.2006.05.010>
- Scholz, U., La Marca, R., Nater, U. M., Aberle, I., Ehlert, U., Hornung, R., Martin, M., & Kliegel, M. (2009). Go no-go performance under psychosocial stress: beneficial effects of implementation intentions. *Neurobiology of Learning and Memory*, 91(1), 89–92, <https://doi.org/10.1016/j.nlm.2008.09.002>
- Schulz-Hardt, S., Rollwage, J., Wanzel, S. K., Frisch, J. U., & Häusser, J. A. (2021). Effects of process and outcome accountability on escalating commitment: A two-study replication. *Journal of Experimental Psychology. Applied*, 27(1), 112 – 124,  
<https://doi.org/10.1037/xap0000321>
- Sharon, I., Drach-Zahavy, A., & Srulovici, E. (2022). The Effect of Outcome vs. Process Accountability-Focus on Performance: A Meta-Analysis. *Frontiers in Psychology*, 13.  
<https://doi.org/10.3389/fpsyg.2022.795117>
- Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes*, 65(1), 1–17. <https://doi.org/10.1006/obhd.1996.0001>
- Simonson, I., & Nye, P. (1992). The effect of accountability on susceptibility to decision errors. *Organizational Behavior and Human Decision Processes*, 51(3), 416–446.  
[https://doi.org/10.1016/0749-5978\(92\)90020-8](https://doi.org/10.1016/0749-5978(92)90020-8)

- Simonson, I., & Staw, B. M. (1992). Deescalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of Applied Psychology*, 77(4), 419–426. <https://doi.org/10.1037/0021-9010.77.4.419>
- Sluga, A. (2021). *Hour of Judgment: On Judgment, Decision Making, and Problem Solving under Accountability* (Nr. EPS-2021-520-F&A). <https://repub.eur.nl/pub/136967/hffinaluploadissertation.pdf>
- Spielberger, C.D., R.L. Gorsuch, R.E. Lushene. (1970). *STAI Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.
- Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, 1(1), 39–60. [https://doi.org/10.1016/0167-2681\(80\)90051-7](https://doi.org/10.1016/0167-2681(80)90051-7)
- Treffenstaedt, C., Brachem, J., & Wiemann, P. (2021). Alfred3 - A library for rapid experiment development (Version 2.3.3b4). Göttingen, Germany: <https://doi.org/10.5281/zenodo.1437219>
- Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4), 1039–1061. <https://doi.org/10.2307/2937956>
- Verwaeren, B., & Nijstad, B. A. (2022). What I do or how I do it – the effect of accountability focus on individual exploration. *European Journal of Work and Organizational Psychology*, 31(3), 421–439. <https://doi.org/10.1080/1359432X.2021.1967322>
- Vieider, F. M. (2009). The effect of accountability on loss aversion. *Acta Psychologica*, 132(1), 96–101. <https://doi.org/10.1016/j.actpsy.2009.05.006>
- Wewers, M. E., & Lowe, N. K. (1990). A critical review of visual analogue scales in the measurement of clinical phenomena. *Research in Nursing & Health*, 13(4), 227–236. <https://doi.org/10.1002/nur.4770130405>

Yates, J. F. (1990). *Judgment and Decision Making* (Bd. 430). Prentice-Hall, Inc.

Zhang, Y. L., & Mittal, V. (2005). Decision difficulty: Effects of procedural and outcome accountability. *The Journal of Consumer Research*, 32(3), 465–472.

<https://doi.org/10.1086/497558>



## Appendix C

### Curriculum Vitae

# Johannes Rollwage

## Curriculum Vitae

### Persönliche Daten

Name: Johannes Rollwage

Geburtsdatum: 27. Juli, 1988

Geburtsort: Salzgitter

Staatsangehörigkeit: Deutsch

### Ausbildung

Georg-August-Universität Göttingen

Seit Oktober 2018: Promotionsstudium im Rahmen des Promotionsprogramm „Behavior and Cognition“ (BeCog) der Georg-August University School of Science (GAUSS)

Georg-August-Universität Göttingen

April 2014 - März 2017: M.A. Soziologie

- Abschlussnote: 1,3 (sehr gut)
- Titel der Masterarbeit: „Normativer Einfluss auf die Ratschlagsnutzung in Judge-Advisor-Systemen“

Georg-August-Universität Göttingen

Oktober 2009 - April 2014: B.A. Soziologie

- Abschlussnote: 2,0 (gut)
- Titel der Bachelorarbeit: „Konstruktion von Männlichkeit und Homophobie im (Leistungs-) Sport“

Michelsenschule Hildesheim

2008: Abitur

- Abschlussnote: 2,4 (gut)
- Leistungskurse: Geschichte, Mathematik & Wirtschaftslehre

Johannes Rollwage

Curriculum Vitae

## Beruflicher Werdegang

Georg-August-Universität Göttingen | Postfach 3744, 37027 Göttingen

Wissenschaftlicher Mitarbeit April 2017 – heute

Studentische Hilfskraft Januar 2017 – April 2017

Ruhrstrat Facility Management GmbH | Am Leinekanal 4, 37073 Göttingen

Separatwachmann Oktober 2012 – Dezember 2016

Baumschule Rollwage | Zur Rast 45, 38271 Baddeckenstedt

Aushilfe Januar 2009 – September 2009

## Publikationen

**Rollwage, J.**, Giersiepen, A., Wanzel, S., & Schulz-Hardt, S. (2020). Der Einfluss bestehender Meinungen auf die Bewertung von Sachinformationen zu Infrastruktur- und Großbauprojekten. *Politische Psychologie*, 8(1), 25-52.

Schulz-Hardt, S., **Rollwage, J.**, Wanzel, S. K., Frisch, J. U., & Häusser, J. A. (2021). Effects of process and outcome accountability on escalating commitment: A two-study replication. *Journal of Experimental Psychology: Applied*; 27(1), 112–124. <https://doi.org/10.1037/xap0000321>

Schulz-Hardt, S., Wanzel, S. K., **Rollwage, J.**, Treffenstädt, C., & Schultze, T. (2021). Do judges prefer advisors with dependent or independent errors? Investigating judges' advice selection and advice weighting. *Journal of Experimental Psychology: General*, 151(7), 1636–1654. <https://doi.org/10.1037/xge0001153>

## Konferenzbeiträge

Rollwage, J., Schulz-Hardt, S., Häusser, J. A. (2022). *Process accountability's potential to improve judgement and decision quality seems to be less robust than previously thought. An attempt to replicate the effect of (process) accountability on loss aversion in 3 studies.* 64. Tagung experimentell arbeitender Psycholog:innen, Köln

Rollwage, J., Häusser, J. A., Schulz-Hardt, S. (2022). *Kann Verlustaversion durch Verantwortlichkeit tatsächlich reduziert werden? Ein Replikationsversuch mittels dreier Experimente.* 52. Kongress der Deutschen Gesellschaft für Psychologie, Hildesheim

## Lehrtätigkeit

Nebenfach: Wirtschaftspsychologie – „Führung“

SoSe 22

- B.A. Soziologie/ B.A. Ethnologie
- 2 SWS

Nebenfach: Sozialpsychologie – „Klassiker der Sozialpsychologie“

SoSe 22

- B.A. Soziologie/ B.A. Ethnologie
- 2 SWS

Wahlpflichtmodul: Sozialpsychologisches Forschen

WiSe 21/22

- B.Sc. Psychologie
- 4 SWS

Nebenfachvertiefung: Wirtschaftspsychologie – „Führung“

SoSe 21

- B.A. Soziologie/ B.A. Ethnologie
- 2 SWS

Nebenfach: Wirtschaftspsychologie – „Führung“

SoSe 20

- B.A. Soziologie/ B.A. Ethnologie
- 2 SWS

## Betreute Abschlussarbeiten

### Bachelorarbeiten:

- Werner, T. (2018). Die Untersuchung des Zusammenhangs von wahrgenommener Führungskräfteprototypikalität und tatsächlichem Führungserfolg.
- Brass, K. (2019). Der Einfluss der wahrgenommenen Prototypikalität einer Führungsperson auf Emergence of Leadership.
- Endres, Y. M. (2019). Follower-Präferenzen für Consideration und Initiating Structure in verschiedenen Ausgangssituationen.
- Güneytepe, E. (2020). Follower-Präferenz bezüglich des Führungsstils (Consideration - Initiating Structure) unter situativen Druckbedingungen.
- Reese, F. (2020). Der Einfluss von Drucksituationen auf Liking und wahrgenommene Qualifikation von mitarbeiter- und aufgabenorientierten Führungskräften Consideration vs. Initiating Structure) durch ihre Follower.
- Klingelhöfer, A. (2021). Einfluss des zeitlichen Horizontes einer Zusammenarbeit auf Führungsstilpräferenzen der Follower.
- Sebon, D. (2021). Effekt des antizipierten Zeithorizontes einer Tätigkeit, auf den präferierten Führungsstil von Follower.

### Masterarbeiten:

- Rurainsky, N. (2020). Follower-Präferenz für Führungsstile und der Einfluss von situativem Druck.
- Bachstädter, J. (2020). Beurteilung von Führungsstilen in aufgabenbezogenen und sozioemotionalen Drucksituationen.
- Ilgner, L. K. (2020). Präferenzen für die Führungsstile Initiating Structure versus Consideration in aufgabenbezogenen und sozio-emotionalen Drucksituationen.
- Böser, J. (2021). Der Einfluss des Zeithorizonts auf die Präferenz von Führungsstilen.