Essays in Behavioral Finance:
Risk, Ambiguity, and Strategic Uncertainty

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum politicarum

an der
Fakultät für Wirtschafts- und Sozialwissenschaften
der Ruprecht-Karls-Universität Heidelberg

vorgelegt von
Christian König genannt Kersting

Heidelberg, Juni 2019

Acknowledgments

While working on my dissertation for the past five years, I have benefited from interacting with many people and I would like to thank all of them. However, there are some people, who deserve special mentioning:

Stefan, thank you very much for taking me as your doctoral student and your extraordinary supervision. Your door has been open on all occasions and you always took the time to discuss our research projects. Your advice on research topics, publication processes, and career planning is invaluable to me. Furthermore, I am especially grateful for the trust placed in me when you sent me off to present first results at the Experimental Finance (EF) conference in Nijmegen not even a year after starting my research; when you allowed me to develop and realize our Masters' seminar on programming experiments; and when we organized EF 2018 and the accompanying Summer School in Heidelberg in 2018. Finally, you offered me long-term employment security, which allowed me to focus on my dissertation without ever having to worry about making ends meet. You have also provided generous funding so I could start my own research projects. Thank you for all of this. I sincerely hope we will continue to work on and start new joint research projects in the future.

Christiane, thank you very much for employing me as your student assistant back in 2012 and giving me the opportunity to be an active part of your research projects. The experience I gained working at your chair made me pursue a doctorate degree and I am very thankful for you to get me in touch with Stefan. Since then, you have always treated me like a regular member of your chair. In addition, I would like to thank you for allowing me to shape our experimental laboratory's operations and for always supporting my ideas for its future development.

Jörg, thank you for agreeing to serve as the second referee for my dissertation and the helpful comments on my thesis.

Markus Pohlmann, thank you for agreeing to serve on my committee without hesitation.

Christoph, for most of the last years, we have shared an office and I would like to thank you for your unconditional willingness to sanity-check my thoughts, exchange research ideas, and discuss politics. Along the way, I learned much more about history than I would ever have on my own. I thoroughly enjoyed our morning coffee breaks and the innumerable amount of jokes and puns we shared.

Martin, Robert, and Liza, together with Christoph you are the very core of our regular lunch group. As such, you have contributed tremendously to the work atmosphere that I have immensely enjoyed. Of course, I have also benefitted incredibly from the constant exchange about our research projects and ideas as well as the comments and corrections you made to my experimental instructions, programs, and papers. Many of these have found their way into the chapters of this thesis. Thank you all.

Luisa, thank you for excellent research assistance. You have supported me in preparing and conducting many of the experiments reported on in this thesis and have saved me hours and hours of work. I am glad to have you as a co-author on one of my papers.

Freya, thank you for your assistance with all of the tedious paperwork as well as any and every bureaucratic hurdle I have encountered. You are the kind soul that connects all members of the chair (and probably even the whole institute).

Sebastian, thank you for many inspiring discussions over the years. You were instrumental in getting me my first job as a student assistant back in Frankfurt. This job got me interested in behavioral economics and ultimately paved the way to my doctoral studies. Thank you for sharing the experiences you made obtaining your own PhD and beyond. I have certainly learned a lot.

My love Carina, not only for enduing my (mostly pointless) complaining about all the things I perceive to be wrong in the world, but also your unconditional support in all aspects of my academic career. Thank you very much for joining me for the adventures to come.

Contents

Chapter 1

Introduction

Uncertainty is a key component of economic decision-making and it can take various forms. In some situations, like gambling, outcomes and associated probabilities are known. Economists commonly refer to this kind of uncertainty as risk. In other situations, the outcomes are known, but their associated probabilities are not. These situations are said to be characterized by ambiguity.[1] However, uncertainty does not have to be the result of individual random events like flipping a coin or rolling dice. There are situations in which outcomes depend on the actions of multiple actors, resulting in strategic uncertainty about the behavior of those involved. How people take decisions under risk, ambiguity, and strategic uncertainty has been studied extensively in a wide variety of contexts. In my dissertation, I focus on the context of financial decision-making and zoom-in on particular situations and mechanisms, which are not fully understood, yet. The thesis consists of two distinct parts. The first part comprises four classic research articles, while the two articles of the second part are rather methodological in nature. I first present a general overview of the two sections, and then introduce each chapter individually.

We start with research on financial decision making for others. Together with my co-authors, I study the factors determining the riskiness of investment decisions that financial advisors make for their clients. Subsequently, we take on the clients' perspective and analyze how they evaluate the investment decisions made on their behalf. In both of these studies, we model investment decisions as decisions under risk. That is, we assume that all possible outcomes as well as the probabilities with which the outcomes occur are perfectly known to the decision maker. In the third article, we take an alternative approach. In this project, we study decision making for others under ambiguity. More precisely, we analyze how people make decisions for others if the outcomes are known, but their probabilities of realization remain unknown. Finally, we broaden the scope even further and acknowledge that in many cases, decision makers do not only face natural uncertainty, but also strategic uncertainty. This chapter reports on coordination games framed in a bank run setting. Specifically, we study how the disclosure of financial information about the fragility of banks and the economic linkages between financial institutions interact.

[1] While there are other definitions and concepts of ambiguity, this is the one we use throughout our articles.

In the second part, we first test the generalizability of existing experimental evidence on countercyclical risk aversion of financial professionals. We conduct the experiment of Cohn et al. (2015) with a student sample to check whether the same results obtain as for financial professionals. Finally, I present software, which I have written to support laboratory managers in more easily managing multiple, parallel installations of the experimental software oTree.

Opening the first part, chapter 2 deals with an important aspect of (retail) financial advice. It concerns the influence financial advisors have over their clients' portfolio composition. Typically, situations of financial advice are modeled as principal-agent relationships in which the principal tasks the agent to make a decision on their behalf. It is commonly assumed that clients asking for financial advice (blindly) trust their advisors and follow their investment recommendations, such that modeling the situation as an agent's decision for the principal is an adequate simplification.

There have been many studies on (financial) decision making for others, which build on variants of this paradigm. A key aspect, however, is overlooked in most of the literature. It concerns the fact that communication takes place between the client and their advisor. One the one hand, a large part of the communication is informal which makes it hard to model and assess systematically. On the other hand, there is also formal communication, which is often mandated by regulators for customer protection purposes. This formal communication typically revolves around the goal of making sure that clients end up with financial products which fit their stage in the life cycle as well as their financial situation and (risk) preferences. The tasks of assessing clients' risk bearing capacities as well as risk and investment preferences fall to advisors, who carry them out as part of Know Your Customer (KYC) efforts mandated by regulators.

Foerster et al. (2017) use data from KYC-forms and advisors' own asset holdings to study how clients' and advisors' preferences interact in shaping clients' investment portfolios. They claim that advisors' risk preferences are the best predictor of the riskiness of their clients' portfolios. Inspired by their research, we take the question of whose preferences determine clients' portfolios to a tightly controlled laboratory setting. The laboratory allows us to improve on several aspects of the existing studies. First, we can rule out confounding effects such as clients deliberately selecting their agents based on their presumed risk preferences. Furthermore, we formalize communication of investment preferences from clients to agents even further and assess how both parties perceive the terminology used. As it turns out, our findings are comparable to those by Foerster et al. (2017) despite the very different approach to studying the question at hand. We find advisors to largely adhere to their clients' explicit investment preferences, yet their own preferences co-determine actual investment levels. In addition, we find evidence for large heterogeneity in the perception of commonly used phrases that describe investment strategies. It is this heterogeneity that can explain why many clients end up with investment levels which do not fit their preferences, despite their advisors' best efforts to do their wishes justice.

Having studied how financial agents make decisions for their clients, chapter 3 concerns the evaluation of the decision by those affected. Again, we model the situation of financial advice by considering an agent who makes a risky investment decision for his principal. In return, the principal has the opportunity to reward his agent for the decision. For example, principal Alice has her fiduciary advisor Bob manage investment decisions for her. If she is content with Bob's investment decisions, she may reward him through various actions. Alice might praise Bob on social media or directly recommend him to friends, maybe she even decides to be more generous in negotiating Bob's fees for years to come. In contrast, Alice might not recommend Bob, demand lower fees or even drop him as an advisor if she believes Bob's performance to be subpar. In this project, we study how these reward decisions are influenced by both, the actual decision and the investment outcome. In many situations, outcome information is the only available signal about the quality of a decision. In these situations, it can be rational to adjust rewards in consonance with observed outcomes. Of greater interest to us, however, are situations in which information about the decision and outcome information is available. Does having knowledge about the outcome influence the evaluation of decision quality in monetarily incentivized situations? Over the course of three experiments, we find strong evidence for outcome bias in the evaluation of the agents, which is robust to monetary incentives, income effects and social preference considerations.

Chapter 4 also concerns decision-making for others. However, we move from decision making under risk to decision making under ambiguity and study if and how attitudes towards ambiguity differ between decisions for one's own account and decisions on somebody else's behalf. We argue that real-world situations, in which financial decisions are taken for another party in the presence or absence of certain accountability conditions, may alternatively be modeled as decision making under ambiguity rather than risk. To this end, we have participants in laboratory experiments take decisions involving ambiguous prospects for their own account, for others and for others with the possibility of being held accountable for their decisions. In similar investment decisions involving *risky* prospects, participants become less risk averse when deciding for others in the absence of accountability, an effect that is mediated by the introduction of accountability (Pollmann et al. 2014). In stark contrast, our study on ambiguous prospects does not reveal similar effects. Attitudes towards ambiguous prospects do not differ between the different decision situations. Predictions of the behavior of financial advisors could be quite different, depending on whether they would be made based on modeling real-life uncertainty as risk or ambiguity.

Broadening the scope from decision making for others, chapter 5 considers situations involving strategic uncertainty. We undertake a project on two crucial elements of the policy debate surrounding financial crises. The first element is risk disclosure. During the financial crises of the 2000s academics and policy makers debated the publication of bank stress test results, which were meant to assess the institutions' individual risk bearing ability. Some have argued that knowledge about banks' fundamentals is important for customers because it enables them to make informed deposit decisions. Critics, on the other hand, have highlighted the potential consequences of disclosing stress test results. They have argued that disclosure may actually trigger bank runs on those banks with the lowest ability to withstand such events. In this context, we ask which level of disclosure (from none to full) has the power to affect depositor behavior in a controlled laboratory experiment adapted from the panic-based bank run model by Diamond and Dybvig (1983). Novel to our approach is the introduction of different levels of precision for information disclosed. Instead of considering only non- and full-disclosure cases, we also study partial disclosure conditions, a necessary and natural extension to the previously existing literature.

The second element is the possibility of financial contagion. Financial contagion occurs if information disclosed about one financial institution leads to a change in behavior of customers of another one. Mostly, this concerns a loss of confidence in the latter based on negative information revealed about the former. Consider the following stylized situation as an example: Information is disclosed about Bank A, revealing a large exposure to the real estate market, which undergoes a major downturn. Depositors run on Bank A in fear of impending bankruptcy. Financial contagion occurs if depositors of

Bank B also withdraw deposits based on the presumption that Bank B has a similar risk exposure to the real estate market. It is noteworthy that no information has been disclosed about Bank B that would justify the reaction. Behavior of depositors may be affected by financial contagion through multiple channels. If the bank run is observable, depositors may actually show herd behavior, without even regarding the information that was initially disclosed. Observing a bank run may also amplify own apprehensions and enable depositors to cross the line from worry to action. However, financial contagion may also take place without observable elements and work purely on the level of beliefs. In our paper, we study this hypothesis as a potential channel of financial contagion. We aim to demonstrate the power of beliefs and assess the economic context in which they excel. To do so, we study conditions which differ in their informational context. We provide participants in laboratory experiments with varying degrees of knowledge about the economic linkages between Bank A and Bank B and subsequently observe how their withdrawal behavior from B changes with the information disclosed about A.

Moving to the methodological second part, chapter 6 presents an extension of a risk taking experiment by Cohn et al. (2015). We ask whether their finding of countercyclical risk aversion among financial professionals is transferable to other samples of the population. Testing the external validity and robustness of research findings is a crucial part of scientific research. Karl Popper famously argued that theories need to be formulated, tested, and, if falsified, replaced with new theories or amended by protective qualifications (Popper 1974). Theories can only be falsified, but never conclusively verified. The more often theories are tested and withstand these falsification efforts, the more likely the theories are true. Extensions and implementations of other researchers' studies in other contexts put the generalizability of their theories and findings to the test. If their theories hold, the extensions add to their substance and credibility. If the original theories do not hold, the exercise still creates knowledge by highlighting the theories' limitations and may provoke a more narrow formulation.

Cohn et al. (2015) find evidence of countercyclical risk aversion in a priming experiment run in a lab-in-the-field setting with financial professionals. Their conclusion, however, suggests that countercyclical risk aversion is a phenomenon, which is not restricted to the specific sample of financial professionals, but might affect all market participants. We put this much more general claim to the test by implementing their priming experiment with students in the laboratory using Cohn et al.'s original materials as far as possible. Our results are surprising. The effects of stock-market-trend priming on participants' willingness to take risks fail to reach statistical significance. Strictly speaking, we are unable to demonstrate that the finding of countercyclical risk aversion in financial professionals extends to the student sample.

Chapter 7 is quite different from the previous chapters in that it presents a software package that I have written to support economic laboratories and especially their managers in providing the infrastructure needed to use Chen et al.'s (2016) experimental software oTree. oTree is written with only a single user in mind. That is, while it can easily be used to run multiple sessions of the same experiment at the same time, it does not support multiple experimenters simultaneously running and adding further experiments or conducting experiments with different language or currency settings in parallel. A commonly suggested solution to these inconveniences is to provide each experimenter with their own installation of oTree. The official manual suggests running multiple instances of oTree on the same computer and carefully adjusting database and network configurations. However, even for experienced laboratory managers, setting up these individual installations manually is a tedious and error-prone effort. The oTree community has also come up with virtual machine managers, which simulate multiple independent computers on a single machine. Each virtual machine is then assigned to an experimenter. A big drawback of this solution is the large resource demand resulting from the overhead of the virtualization of complete systems.

The software I have written and open sourced uses relatively novel software distribution and virtualization techniques. Similar to the community efforts, my software creates virtualized environments, which can be individually assigned to experimenters. However, building on the Docker platform, my solution is much less resource demanding and thus allows for more parallel users on the same hardware or, alternatively, the use of less powerful machines for the same number of experimenters. My software also provides an intuitive web interface, which allows new installations to be created with only a few clicks. Finally, it also makes life easier for experimenters, because frequently used features can be accessed from the web interface and do not require cumbersome command line interaction.

The following chapters contain the individual research articles. Some of them are immediately followed by short appendices. Further supplementary material is available online with each chapter referencing the relevant resources in a footnote. Chapter 8 concludes my thesis.

Chapter 2

Investing for Others: Principals' vs. Agents' Preferences

Abstract. We study the degree to which financial investment advice is driven by the client's preferences, versus the preferences and incentives of the advisor. In a typical financial advice setting, clients can communicate their preferred investment profile to their financial advisor. We observe a high willingness of advisors to follow their clients' preferred investment profiles, but also replicate evidence that advisor preferences matter for investment choices. However, even though advisors are willing to follow their clients' preferences, they often fail to do so from their clients' perspective. One reason is that people are very heterogeneous in their perception of the terminology commonly used to describe the riskiness of financial investments.[2]

[2] This chapter is co-authored by Luisa Kling and Stefan T. Trautmann.

2.1 Introduction

As part of the revised Markets in Financial Instruments Directive (MiFID II) financial advisors in the European Union are obliged to assess their customers' personal attitudes towards taking risks, their risk tolerance, and their risk bearing capacity (Hallahan et al. 2004). Similarly, investment advisors in the United States face a *duty to inquire* and a *duty to give only suitable advice,* which entail assessments of the risk tolerance and risk bearing capacity of their clients. Clearly, these are neither easy nor clearly defined tasks and their implementation varies widely ranging from customer risk attitude questionnaires to behavioral measures of risk preferences (Grable and Lytton 1999; Kaufmann et al. 2013, Roszkowski and Grable 2005). Independent of jurisdiction, the goals of these regulatory efforts are to align the interests of clients and advisors and prevent the former from fraudulent exploitation by the latter.

One of the single most important questions for all stakeholders in situations of financial advice is how financial advisors shape their clients' investment portfolios. For clients, it is a question of optimal life cycle asset allocation. If advisor characteristics affect their portfolio allocations, advisor selection becomes a variable in their optimization. For advisors, own financial interests and ethics play a major role. For regulators, finally, consumer protection as well as welfare considerations are key. Foerster et al. (2017) study the question using data from Canadian mutual fund dealers. Their data contains both stated investment preferences from Know Your Customer (KYC) forms and actual investment portfolio holdings. They find that advisors' own risk attitudes are the strongest predictor for the risky investments on behalf of their clients. Their results show that customization of portfolios to match different customers' needs is limited. Despite the richness of the empirical datasets, the authors lose control compared to studies based on laboratory experiments. Specifically, it remains unclear how matching between advisors and clients affects the results. Clients might specifically select advisors based on a number of different and potentially unobservable characteristics. Similarly, it might be the case that advisors simply use their own risk tolerance as their best predictor for clients' risk tolerance if the communication of risk preferences from clients to advisors (via KYC forms) is too unspecific.

The goal of this paper is to revisit the question of how financial advisors shape their clients' portfolio in a tightly controlled laboratory setting with randomized treatment and role allocations. We elicit participants' perceptions of common investment profile terminology used in financial advice, let clients communicate their preferred profile to their advisor, and observe the advisors' subsequent investment decisions. This lets us test whether the strong effect of advisors' own preferences reported by Foerster et al. (2017) survive in a more tightly controlled setting absent the possibility of selection. We also ask whether customization of client portfolios takes place and how different compensation schemes affect advisors' decisions. Finally, we study the effects of ambiguous communication of risk preferences on investment decisions.

In a 2-by-3 between-subject design, advisors either take a decision for only one client or for a group of five clients and receive either a fixed payment or earn a share of the profit or the client's outcome. In the first part of our experiment, participants individually and privately map a set of investment profiles, which range in wording from "very conservative" to "aggressive growth", to investment shares into a risky asset. The terms used to describe the investment profiles are commonly used in financial advisory documents (Mutual Fund Dealers Association of Canada 2014, subsequently MFDA). In the second part, participants take a Gneezy and Potters (1997) investment decision: Clients choose one of the five investment profiles which is subsequently communicated to their financial agent. Knowing their clients' preferred strategies, financial agents then decide how much of their clients' endowments to invest in the risky asset.

Initially collecting the individual mappings of investment profiles into risky investment shares offers us the unique opportunity to investigate the perception of the risk profiles. We find a sizable overlap of the investment profiles and conclude that the perception of risk attitude terms commonly used in financial advice is very heterogeneous. Hinting at this issue, Bradbury et al. (2015) emphasize the importance of understanding the risks involved in investment decisions and show that these can be improved by simulating experience compared to survey-style risk assessment procedures. Further adding to the evidence, Glaser et al. (2019) demonstrate that risk perception concerning financial assets is sensitive to the presentation format. While we do not systematically vary the presentation format, we are still able to control for the perception of the investment terms in our analyses and identify to which degree mismatches in invested amounts and investment preferences can be traced back to differences in perceptions between advisors and clients.

We carefully examine the behavior of advisors given their own perception of the investment profiles and find that they invest in a way that is compatible with their clients' investment profile preferences in almost half of all cases. Observations from our Group treatment reveal that tailoring of investments to clients' preferences does not only occur on the aggregate, but also on the individual level. Yet, advisors' own investment preferences also affect their clients' portfolio. Taken together, we qualitatively replicate the findings by Foerster et al. (2017), although the effect of advisors' own preferences seems to be much weaker in our tightly controlled laboratory environment compared to the empirical real-world data.

Turning to the effects of different compensation schemes on the behavior of advisors, we find that the degree to which they follow clients' stated preferences is hardly affected by them. This observation is consistent with previous evidence provided by Ifcher and Zarghamee (2018), who find that agents have a tendency to act as surrogates for their principals. Even with strong financial incentives for the advisors to disregard their clients' preferences, the clients' preferences still substantially determine the level of investments in their experiment. The observation that agents' financial motives do not affect their behavior much is corroborated by Rud et al. (2019), who show that financial incentives do not increase misreporting of agents to clients in their study of different market structures.

Next, we take an outcome perspective and ask whether clients get "what they want". We find evidence of a substantial problem of communication between advisors and clients: That is, even though advisors are keen to follow their clients' preferences and actually do so according to their own perception of the investment profiles, they often do not succeed from their clients' perspective, simply because they differ in their understanding of the investment profile. Clients end up with investment levels which are incompatible with their preferences, despite the advisors' attempts to align the two.

Finally, we consider two control conditions. In the first, we remove uncertainty about the perception of the different investment strategies. This condition is aimed at removing the fundamental translation error between clients' and advisors' understanding of the investment strategies. In the second control condition, we remove accountability and frame the experiment neutrally, instead of in a financial decision making context. This condition allows us to assess to what degree the possibility of holding advisors accountable for their actions contributes to the large proportion of advisors following their clients' preferences.

The remainder of the article is organized as follows: In the next section, we present a short overview of the existing literature on risk taking for others. In section 2.3, we present our experimental design and the procedures. Section 2.4 shows the results and section 2.5 provides a short discussion of the results. Section 2.6 concludes.

2.2 Related Literature

A growing body of literature on risky decision making for others is focused on determining whether risky decisions for others are different from risky decisions for oneself. If a difference exists, the question of the direction emerges: Do advisors take higher or lower levels of risk for their clients than they do for themselves? The evidence is mixed. This section provides a short overview of the existing literature. We start by providing some evidence for advisors taking higher levels of risk when deciding for others.[3]

Pollmann et al. (2014) employ the Gneezy and Potters (1997) investment task with agents taking decisions for one principal. Comparing their decisions to agents who decide for themselves, they find them taking less risk averse investments when deciding for others. Furthermore, Andersson et al. (2014) use a multiple price list method to study risk taking for others both in situations when losses are possible and when they are not. They do not find any difference in risk levels taken between decisions for themselves and for others if losses are impossible. Still, participants' decisions involve more risk when deciding for others if losses are possible. Another finding is that higher levels of risk taking are primarily driven by a decrease in loss aversion. Hence, the authors conclude that making decisions for others has a de-biasing advantage over decisions for oneself. This is in line with the findings of Polman (2012). He shows the stable result in several studies that decisions for others involve less loss aversion than decisions for oneself. Moreover, Pahlke et al. (2015) study the effect of being responsible for someone else's payoff on risk attitudes. In the gain domain, they find an increase in risk aversion. However, in the loss domain, they observe more risk seeking behavior. Due to their finding of an increase in risk seeking under responsibility for small probabilities in the

[3] As some studies measure risk seeking behavior and others measure risk aversion, we report both studies which find higher levels of risk seeking as well as lower levels of risk aversion in the subsequent paragraphs.

gain domain, they reject the hypothesis of a cautious shift when being responsible for other peoples' payoff.

By using both a multiple price list experiment as well as a first-price sealed-bid auction, Chakravarty et al. (2011) find that subjects are less risk averse when deciding for others as compared to deciding for themselves. Further, they apply a belief elicitation method to get to the finding that people do not try to act in accordance with what they believe are the risk attitudes of their principals. Hsee and Weber (1997) investigate how people predict the risk preferences of others and examine possible mechanisms that people may use when estimating others' risk tolerance. They find evidence for the *Risk-as-Feeling*s hypothesis according to which "people predict others to have similar risk preferences to themselves, but they predict others to be more risk neutral than themselves" (Hsee and Weber 1997, p. 45). According to this hypothesis, people base their predictions of other peoples' risk preferences both on their own feelings towards risk as well as on risk neutrality because they have problems in imagining that people have feelings that are as strong as their own. Hereby the extent to which people base it on their own feelings depends on how vivid the other person is. Thus, when the other person is abstract, they base their predictions to a larger part on risk neutrality and hence overestimate others' willingness to take risks.[4]

Besides the findings of increased risk taking in decisions for others, there is also some evidence for lower levels of risk. First, Reynolds et al. (2009) compare decisions of participants when they decide between a safe and a risky option for their own and when they decide between the same options for a group of people. They find them choosing higher levels of risk when deciding for themselves as compared to deciding for others. Eriksen and Kvaløy (2010) find that participants take significantly lower levels of risk when they make investments for other people as compared to making investments for themselves. The authors interpret this finding by means of the empathy gap (Loewenstein 1996) such that agents underestimate their principals' willingness to take risks. In Charness and Jackson (2009), participants play a stag-hunt game. In one treatment, they take the decision for their own account, while in the other treatment a participant takes the decision for another passive participant as well. They find less subjects choosing the risky option when another player earns the same payoff. Montinari

[4] The term abstract refers to not seeing that person or having a picture of her.

and Rancan (2018) use lotteries with negative expected returns. They find participants investing more for themselves than for friends. Yet, they find no difference in investments for themselves and on behalf of a stranger. Bolton and Ockenfels (2010) let participants choose between a risky and a safe option. They compare the decisions when they affect the chooser's payoff only and when they affect both the chooser's as well as another participant's payoff and find that choices are more risk averse in the latter situation.[5] Füllbrunn and Luhan (2017) hold the variety of different designs responsible for the different results. They point out differences concerning the payoff alignment between agents and principals in the existing literature. On the one hand, agents take decisions for their principals only and earn a fixed payment.[6] On the other hand, the same decision is implemented for themselves.[7] In their own experiment, they find evidence for a cautious shift, which is independent of payoff alignment. Additionally, they find that agents invest according to what they believe their principals wish to invest for themselves, which stands in contrast to Chakravarty et al. (2011).

By means of our experimental design, we aim to address this controversy. First, we give principals the opportunity to communicate their preferred investment profile to their agent, thereby reducing the information asymmetry. Furthermore, since we know the agents' perception of the investment profiles, we can distinguish two reasons why mismatches happen: Either the agent deliberately chooses not to follow the principals preferred profile or he perceives the profile differently and follows the principals request according to his own notion of the terms.

2.3 Experimental Design

2.3.1 Overview

During the course of the computerized laboratory experiment, participants pass three stages and take on both the role of a client and a financial advisor. The experiment starts with the Profile Perception Stage, in which participants are asked to map investment profiles onto an investment scale ranging from 0% to 100%. In the Preference Stage, we elicit participants' own investment preferences as a client. Finally, we put them into the

---

[5] This holds as long as choosing the safe option does not imply inequality to the detriment of the chooser.

[6] Andersson et al. 2014, Chakravarty et al. 2011, Eriksen and Kvaløy 2010, Montinari and Rancan 2018, Pollmann et al. 2014, Polman 2012, and Reynolds et al. 2009.

[7] Andersson et al. 2014, Bolton and Ockenfels 2010, Charness and Jackson 2009, and Pahlke et al. 2015.

roles of financial advisors to take an investment decision for other participants. In this Investment Stage, financial advisors are informed about their clients' investment preferences before making their decision. The experiment concludes with a short demographics questionnaire.[8]

2.3.2 Investment Profile Perception

In the Profile Perception Stage, we present participants with investment profile names, which are commonly used in the financial industry.[9] Participants learn that there are two investment opportunities: a safe and a risky asset. We then ask each participant to map the investment profiles into ranges of investment amounts in the risky asset on a scale from 0% to 100%. That is, we ask participants to reveal which levels of investment into a risky asset they think of when confronted with each investment profile. We enforce consistency, i.e. that investment profiles which imply greater risk appetite than others cannot be mapped into lower risky investment levels. The Profile Perception Stage provides us with an individual measure of how participants perceive the investment profiles in terms of the investment ranges in the Gneezy and Potters (1997) setup.

Figure 2.1, Panel A shows the starting point of the mapping procedure as it was presented to the participants on their screens. Starting with the investment profile *very conservative* participants can successively drag and drop each profile box onto the scale. Participants can adjust the size of each box, i.e. adjust lower and upper limits of an investment amount in the risky asset such that it matches their perception of the investment profile. Panel B shows an example of an intermediate step in the elicitation process. In this example, the participant has already mapped two of the profiles to risky investment levels and has adjusted the ranges they cover. Panel C finally shows an example of the completed elicitation process. The participant perceives a risky asset share of roughly 0-10% to match a v*ery conservative* profile. The *conservative income* profile covers a wide range of risky asset shares from approximately 10% to 50%. A risky asset share of 50-70% maps into a *balanced* profile. Finally, 70% to 80% and 80% to 100% are considered adequate for *growth* and *aggressive growth* profiles, respectively. Note that the full range of 0% to 100% had to be covered by the five profiles. Simply dragging them onto the scale was not enough, as they would only cover about 80% of the range by

[8] Data sets, instructions and further supplementary material are available online at https://ckgk.de/files/thesis/ch2_investing_for_others.zip.

[9] These are used by the Mutual Fund Dealers Association of Canada (2014).

default. Participants had to adjust the size of at least one profile to be able to continue. This was implemented to make sure participants had to familiarize themselves with the range adjustment feature.

Figure 2.1: Investment Profile Perception Elicitation

Notes: The figure shows the process of the investment profile perception elicitation. Panel A shows the starting point of the mapping procedure as it was presented to the participants on their screens. Panel B shows an example of an intermediate step in the elicitation process. In this example, the participant has already mapped two of the profiles to risky investment levels and has adjusted the ranges they cover. Panel C finally shows an example of the completed elicitation process. Note that the full range of 0% to 100% had to be covered by the five profiles. An animated version is available at https://youtu.be/mcTX1QQX2f4.

At this point of the experiment, participants only know that there will be a risky and a safe asset. We consciously forgo a more detailed description of the assets in order to better resemble the situation in an actual financial advice setting. It is important that risk assessment tasks are free of complex details to foster understanding (MFDA 2014). Precise details of the financial products are typically only provided to clients at a later stage of the process, when the actual product selection takes place. In the preceding assessment stages, products are commonly abstracted away from and portfolio composition is presented in a simplified manner. They focus, for example, on the broad categories of equity and fixed income assets only (cf. sample investor profiles and asset allocations in MFDA 2014).

2.3.3 Investment Preferences

In the Preference Stage, we make participants familiar with the details of the Gneezy and Potters (1997) investment task in the agency setting: The client owns an endowment of

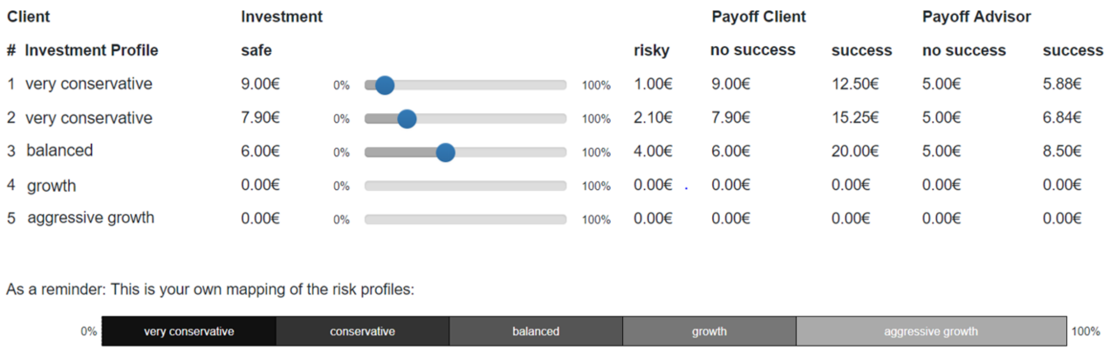
10 Euro, which the advisor has to allocate between a safe and a risky asset. The risky asset resembles a lottery and has a return of $+250\%$ with probability $p = 1/3$ and a return of -100% with a probability of $1 - p = 2/3$. The expected return of the risky asset is 16.67%. The safe asset has a return of 0%. The advisor decides to invest an amount $x \in [0,10]$ in the risky asset. The remainder $10 - x$ is automatically put into the safe asset. In this stage, all participants take on the roles of clients and state their investment preference by selecting one of the investment profiles they already encountered in the Profile Perception Stage. The selected profile is then communicated to the advisor in the Investment Stage. Participants are reminded that the preferred profile is communicated with the intention that the advisor uses the information when making the investment decision. While this rather explicit demand for compliance with the clients' preferences might seem unconventional for a typical laboratory experiment, it is a very natural aspect in the context of financial advice. Clearly, all of the communication between clients and advisors is aimed at informing and guiding the advisors' subsequent actions in real-life situations. This is especially true if communication takes the form of an investment preference assessment initiated by the advisor.

2.3.4 Investment Decisions

Finally, in the Investment Stage, all participants become financial advisors and make the investment decision for their clients. In this stage, advisors are informed about the investment profile selected by their clients in the Preference Stage. Advisors are not bound by their clients' investment profile preference, but can freely choose any feasible investment in the risky asset. When deciding on how much to invest on their clients' behalf, advisors have full information: For each client they see the preferred investment profile. For reference, they are also reminded of their own mapping of investment profiles into investment levels in the risky asset. Advisors make their investment decision by moving a slider to set the risky investment for each one of their clients. Next to the slider, advisors see the clients' resulting minimum and maximum payoffs as well as their own resulting minimum and maximum advisor payoffs. The payoff displays update with every move of a slider for instant feedback on the effects of different investment levels. Advisors always take the investment decisions for all of their clients simultaneously on the same screen. This allows them to easily differentiate investments between different profile preferences, if they intend to do so. Figure 2.2 shows an example of the decision screen.

At this point, agents and clients are also aware of a weak accountability mechanism: After learning about their payoff relevant role, the investment decision of their advisor and their final payoff, clients are asked to send a short message to their advisors expressing their (dis)satisfaction with the investment decision. The pre-defined messages read “I am [very satisfied / satisfied / dissatisfied / very dissatisfied] with your decision”.

Figure 2.2: Agents’ Decision Screens



Notes: The figure shows the lower half of the advisors’ decision screen in the Group treatments. The first column shows the investment profile communicated by each of the five clients. The next three columns show investments in the safe and risky assets as well as the decision slider, which is used to allocate the endowment between the two. In this example, the decision maker has already set investments for the first three clients, but has not started to select investments for the last two (no default slider position). The next two columns show the payoffs the clients receive in the investment success / no success cases. The final two columns show the corresponding payoffs to the advisor. All values in the table update instantly with slider movements. Below the decision table, a reminder of the agent’s own mapping of the investment profiles to investment shares in the risky asset is shown. An animated version is available at https://youtu.be/s7IS2FRWY1o.

2.3.5 Treatments

Using a 2-by-3 between-subject design, we systematically vary the number of clients on whose behalf advisors have to take the investment decision as well as the payment schemes for advisors. In the Single treatments, advisors take the investment decision for exactly one client whereas in the Group treatments, advisors take the decision for a total of five clients simultaneously. Advisors can set the investment for each of their five clients individually. In the Fixed payment scheme, advisors get a fixed payment for their investment decision. Under Limited Liability, advisors get a fixed payment plus an additional share of the positive return of the investment decision. That is, they do not face any downside risk. Finally, in the Co-Investment condition, advisors get a fixed payment and a share of their client’s portfolio after the investment decision and its outcome have materialized.

Single and Group Treatments

In the Single treatments, the computer matches two participants within a session. We are particularly interested in situations in which a client's and an advisor's preferred investment strategies differ. Therefore, we match them such that we observe the highest possible variability of investment preferences within pairs. Unbeknownst to them, both participants take the investment decision as advisors for each other. After all investment decisions have been made, one of the two participants in a pair is randomly selected to be the payoff-relevant advisor, the other one becomes the client.

In the Group treatments, participants are allocated into groups of six. We introduce this treatment in order to increase the probability of agents observing heterogeneous investment preferences of their clients and hence being able to observe the extent to which they differ. We again match groups to maximize the variability of preferred investment profiles. Every participant takes the investment decision as an advisor for every one of the five other participants in the group. Finally, we randomly select three participants of each group to be the payoff-relevant advisors and randomly match each one of them with one of the remaining three participants, who become clients. We choose three advisors from each group in order to keep the probability of being an advisor constant across treatments. Thus, participants in both the Single and Group conditions face a 50% probability of being paid according to their decisions as financial advisors.

The group size of six participants is motivated by our desire to expose participants to the largest possible variation in investment strategies preferred by their agents. With a group size of six, each participants takes the decision for five clients, which is exactly the number of available investment profiles. Yet, only 4 out of the 108 Group treatment participants faced the maximum variety and observed five different investment profiles. 53 of the participants saw four different investment profiles and 47 encountered three different ones. Four participants only observed two different profiles and there was no case in which participants faced just one profile. In total, 96.3% of our participants saw at least three different investment profiles and were thus exposed to a reasonable degree of heterogeneity.

Payment Schemes

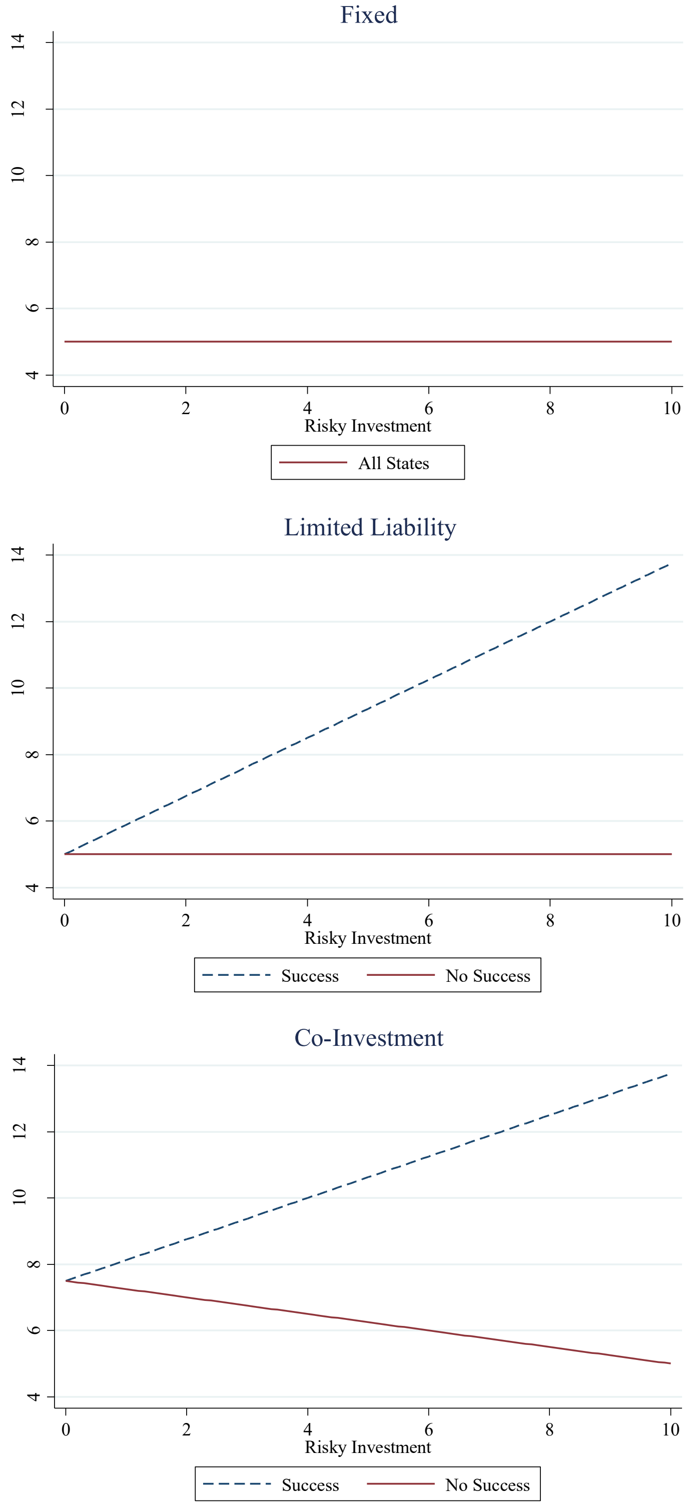
We further systematically vary three payment schemes put in place for the financial advisors. Under all payment schemes, clients are paid according to the investment task. In the Fixed payment, scheme advisors get a fixed payment of 5 Euro whereas in the Co-

Investment and Limited Liability payment scheme, advisors' pay is partially linked to their investment decision(s). Under the Limited Liability compensation scheme, advisors receive a fixed payment of 5 Euro plus a share of 35% on the positive return of their corresponding clients. That is, advisors do not face any downside risk, because their compensation is bounded below by the fixed payment, which is independent from investment success. However, they do have clear and substantial risk taking incentives to increase their own payoffs, creating a situation of limited liability.

The Co-Investment compensation scheme lies in between the two extremes. Under this compensation scheme, advisors receive a fixed payment of 5 Euro plus a share of 25% on the payoff of their corresponding clients. In contrast to the Limited Liability treatment, advisors' face a downside risk because they can also lose by choosing riskier investments. Still, advisors' expected earnings increase as they invest more in the risky asset. That is, advisors face a similar payoff structure as their clients but in an attenuated form: The variance in payoffs is lower compared to their clients' and in worst case, they end up with a payoff of 5 Euro whereas their clients can end up with a payoff of zero.

To simplify the experiment, advisors' compensations are always paid by the experimenter and do not come out of clients' portfolios. Figure 2.3 shows the advisors' earnings as a function of the investment in the risky asset for our payment schemes.

Figure 2.3: Advisors' Compensation Schemes

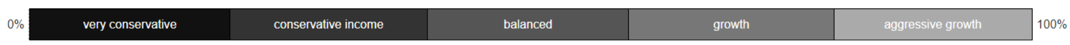


Notes: The figure shows the three payment schemes put in place for the financial advisors. In the Fixed payment scheme advisors get a fixed payment of 5 Euro. Under the Limited Liability compensation scheme, advisors receive a fixed payment of 5 Euro plus a share of 35% on the positive return of their corresponding clients while under the Co-Investment compensation scheme, advisors receive a fixed payment of 5 Euro plus a share of 25% on the payoff of their corresponding clients.

Additional Control Treatments

We also conduct two additional control treatments. The first aims at examining how the uncertainty surrounding the understanding of the investment profiles affects the decisions. Thus, in the *Certainty* treatment, we modify the profile perception stage, while all other stages stay unchanged. In contrast to our main treatments, we do not elicit participants' perception of each investment strategy. We rather establish a common understanding of these terms. This is done by showing participants the five investment profiles and explicitly defining how they are supposed to map into different investment levels.[10] Each investment profile now covers a fixed range of 20% as shown in Figure 2.4. Fixing the perception of the profiles removes the possibility of observing unintended mismatches: If an advisor follows his client's preferred profile, the client will perceive the advisor's behavior as in line with his investment request by design. If there is a mismatch, it must be because of advisors deliberately choosing an investment that is incompatible with clients' preferences. The remaining experiment stays unchanged: Clients pick their preferred investment profile, which is communicated to their advisor. Advisors make the investment decisions. The compensation is analogous to the Limited Liability treatment. We only run the Single variant of our design for the Certainty condition.

Figure 2.4: Preference Perception Stage



Notes: In the Certainty treatment, we establish a common understanding of the investment strategies by fixing each interval to a size of 20%.

Note that our main treatments all include accountability aspects, which might be driving the effects we observe: 1) the experiment is framed in a finance context; 2) clients can tell their advisors how they would like them to invest; and 3) clients can send messages expressing their satisfaction or dissatisfaction with their advisors' decision after they learn about the investment decision and its outcome. Thus, in a second control condition (*No Accountability*), we remove these aspects. The instructions are neutrally framed[11], there is no elicitation and no explicit communication of investment preferences, and clients can no longer express their satisfaction or dissatisfaction with the advisors'

[10] We make sure participants engage with the scale and understand it correctly by asking additional comprehension questions in this treatment.

[11] For example, we use "decision maker" and "recipient" instead of "advisor" and "client".

decisions. In line with the first control condition, we again run the Single / Limited Liability variant only.

2.3.6 Procedures

The experiment was conducted in the experimental laboratory at Heidelberg University in Germany. Sessions were organized with the software hroot (Bock et al. 2014) and the experiment was programmed using oTree (Chen et al. 2016). Participants entered the laboratory and were randomly placed at one of the 20 separated computers. All instructions were displayed on-screen and questions were answered in private. We ensure understanding of the instructions by letting participants advance through the instruction section only after answering a set of comprehension questions correctly. The experiment concluded with a short demographic questionnaire. Participants received cash payments in private and were dismissed from the laboratory. A total of 434 participants took part in the experiment (56.2% female, 30.2% economics students, average age: 23.0). In total, we ran 26 sessions (6x 3 for the main treatments and 2x 4 for the additional control conditions) with 324 participants in the main treatments and 110 participants in the controls. Each session lasted about 45 minutes and participants earned an average amount of 11.85 Euro including a show up fee of 4 Euro.

2.4 Results

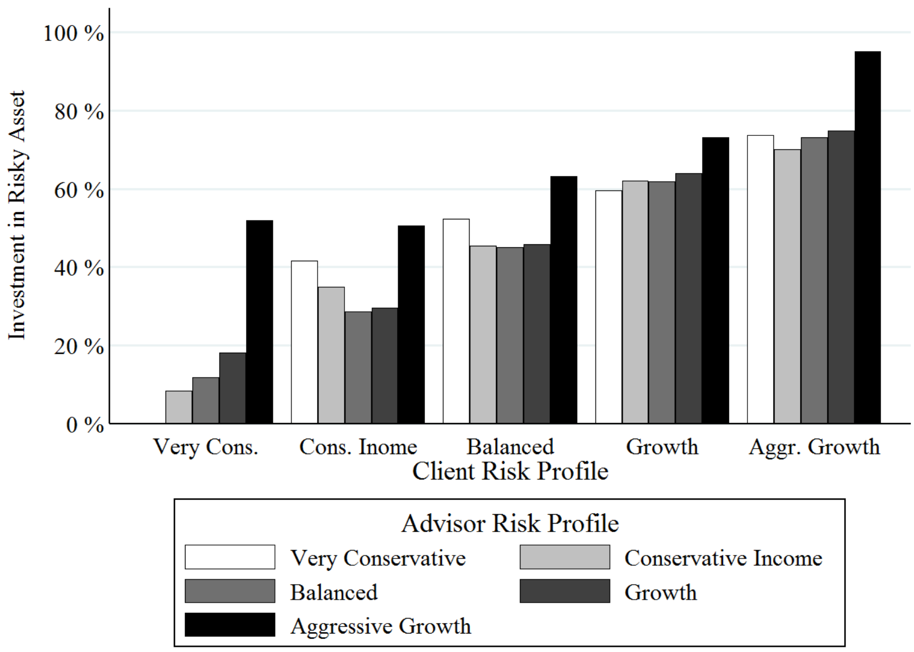
Our main intention is to investigate what drives risky investment shares in an agency setting. To do so, we divide the analysis into two subparts. We focus on advisors' behavior first and investigate whether they follow their clients' profiles or rather base their decision on their own risk preference. Next, we take on the perspective of clients and investigate whether they "get what they want". As an intermediate step, we examine the perception of the investment profiles and how differences thereof might affect the decisions taken. Unless otherwise stated, we base the results on our six main treatment conditions. We only draw upon the data from our control conditions in the discussions in section five.

2.4.1 Advisors' Behavior

Visual Inspection

We start our analysis by an examination of the investments in the risky asset. We are interested in whether advisors follow their clients' preferred investment profiles or if they implement investments that correspond to their own risk preference. Figure 2.5 shows the average investment in the risky asset for different combinations for the clients' and advisors' preferred investment profiles. In line with Foerster et al. (2017), we find that advisors' own risk preferences influence the risk they take on behalf of their clients. Within each profile preferred by clients, we find that the average investment in the risky asset increases with the preferred profile of the advisor. A first visual inspection reveals that both the risk preference of the client as well as the risk preference of the advisor seem to play a role when taking risky decisions on behalf of others.

Figure 2.5: Investment in the Risky Asset by Clients' and Advisors' Profiles



Notes: This figure shows the average investments in the risky asset for each client and advisor profile combination. Client Risk Profile refers to the preferred investment profile of the client while Advisor Risk Profile refers to the preferred investment profile of an advisor.

As a second step, we are interested in whether advisors follow their clients' investment profile given how they perceive the scale of investments in the risky assets and the profile of their client. That is, we base the analysis in this section on whether advisors implement the profile of their clients according to the advisors' perception, irrespective of how the client himself perceives the profile. Indeed, in 49.3% of the decisions over all main treatments, advisors follow their clients' preferred investment profile.[12] This is despite the fact that none of our payment schemes provides incentives

[12] If we allow for a 'wiggle room' of 5 percentage points (0.50€ in the investment task) for the perception of the profiles, advisors follow their clients in 59.9% of the decisions.

to follow the clients' wishes. In contrary, the Limited Liability conditions even unambiguously incentivizes advisors to take risks above and beyond their client's preferences for own monetary gain.

Table 2.1: Risky Investment Shares by Treatment Condition

	Compensation		
	Fixed	Co-Investment	Limited Liability
Single	47.8%	50.7%	50.9%
Group	46.9%	50.1%	56.5%

Notes: For treatment Single the number of observations is 54 for each compensation treatment. For Group it is 270, because we observe five investments decisions (not independent) for each participant.

Risky Investment Shares

Table 2.1 provides an overview of risky investment shares separated by treatment conditions. In order to investigate advisors' investment behavior more formally and test for treatment differences, we use OLS regressions to estimate the investment in the risky asset. In specification (1), we regress the risky share on the advisors' and the clients' preferred investment profiles, representing their risk preferences. In specification (2), we add treatment indicators and their interactions, as well as control variables. Table 2.2 reports the results. Disregarding treatment differences, clients' and advisors' preferred investment strategies already explain a large fraction of the observed variation. The effect of clients' preference on the amount invested into the risky asset is larger than the effect of advisors' preferences (F-test, p-value < 0.01 for specifications (1) and (2)). When considering our treatment conditions, we observe that investments are lower in the group conditions under fixed payments, but react differentially to the two other compensation schemes. We therefore conclude that advisors base their investment decisions to large parts on their clients' preferences but also consider their own risk preferences. This is generally in line with the visual impression of Figure 2.5.

Table 2.2: Regression Analysis Investments in the Risky Asset

	Investment in Risky Asset (1)	Investment in Risky Asset (2)
CI Treatment		0.02
		(0.35)
LL Treatment		-0.05
		(0.35)
Group Treatment		-0.44*
		(0.26)
CI × Group		0.92**
		(0.45)
LL × Group		0.88*
		(0.49)
Profile Client	1.49***	1.52***
	(0.07)	(0.07)
Profile Advisor	0.28**	0.30**
	(0.13)	(0.13)
Age		0.05
		(0.03)
Female		0.01
		(0.24)
Constant	-0.60	-2.07**
	(0.47)	(0.87)
Observations	972	972
R^2	0.37	0.40

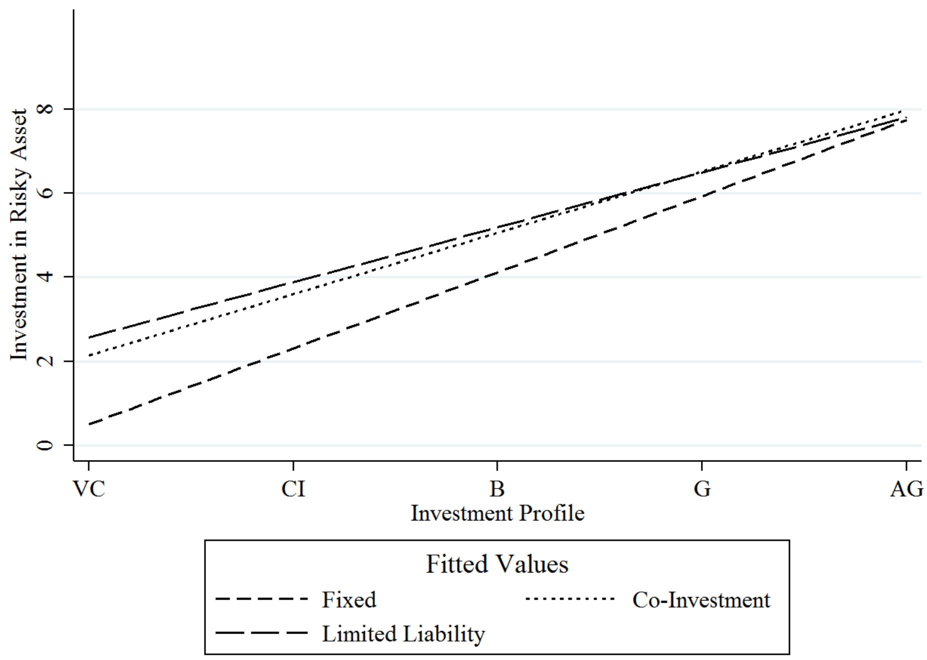
Notes: We report OLS regression coefficient estimates with robust standard errors in parentheses. The standard errors are clustered on the individual level. The dependent variable is investment in the risky asset. CI and LL indicate the treatment conditions Co-investment and Limited Liability, respectively. Profile Client indicates the category the client has chosen as preferred investment strategy. Profile Advisor indicates the profile the advisor has chosen as preferred investment strategy in the Preference Stage. ***/**/* indicate significance at 1%/5%/10%.

Portfolio Customization and Monetary Incentives

While we observe that about half of our advisors factually do not invest in a way that is in line with their clients' preferences, they might still have the intent to do so, but fail in implementing their intent. The group treatment makes the heterogeneity of different investment profiles among an advisor's clients salient. The advisors in this condition are aware that clients have different tastes. By measuring how strongly individual advisors differentiate between clients with different investment profile preferences, we can uncover the advisors' intentions to follow their clients' preferences. The more they take their clients into account, the stronger they should differentiate investments between profiles. The less importance they put on clients' preferences, the more similar should be the invested amounts for all clients. Furthermore, we are interested in whether the compensation scheme affects the extent of differentiation between clients with different investment preferences.

Due to the monetary incentives under the Limited Liability compensation, we expect advisors to invest more and differentiate less as compared to the Fixed treatment. Figure 2.6 shows the differentiation of agents' investments for their clients for our three compensation schemes (Appendix 2A shows the differentiation of individual advisors by comensation treatments). The degree of differentiation is highest under the Fixed compensation and lowest under the Limited Liability compensation. The correlation coefficients are all positive and significantly different from zero (Fixed: $\rho = 0.79$, $p < 0.01$; Limited Liability: $\rho = 0.49$, $p < 0.01$; Co-Investment: $\rho = 0.61$, $p < 0.01$, all of them are spearman correlation coefficients). The correlation between the clients' profiles and the investment in the risky asset is strongest under Fixed compensation and (marginally) significantly different from both correlations under Limited Liability (0.79 vs. 0.49 , $p < 0.01$) and Co-Investment compensation (0.79 vs. 0.61, $p = 0.055$). That is, we find high levels of customization of investments for clients. Yet, even under the strongest of financial incentives, advisors do not disregard their clients' preferences.

Figure 2.6: Risky Investments in Group Treatments by Compensation Scheme



Notes: The graph shows aggregated investments for each communicated investment profile in the Group conditions. We plot separately fitted values for each compensation scheme.

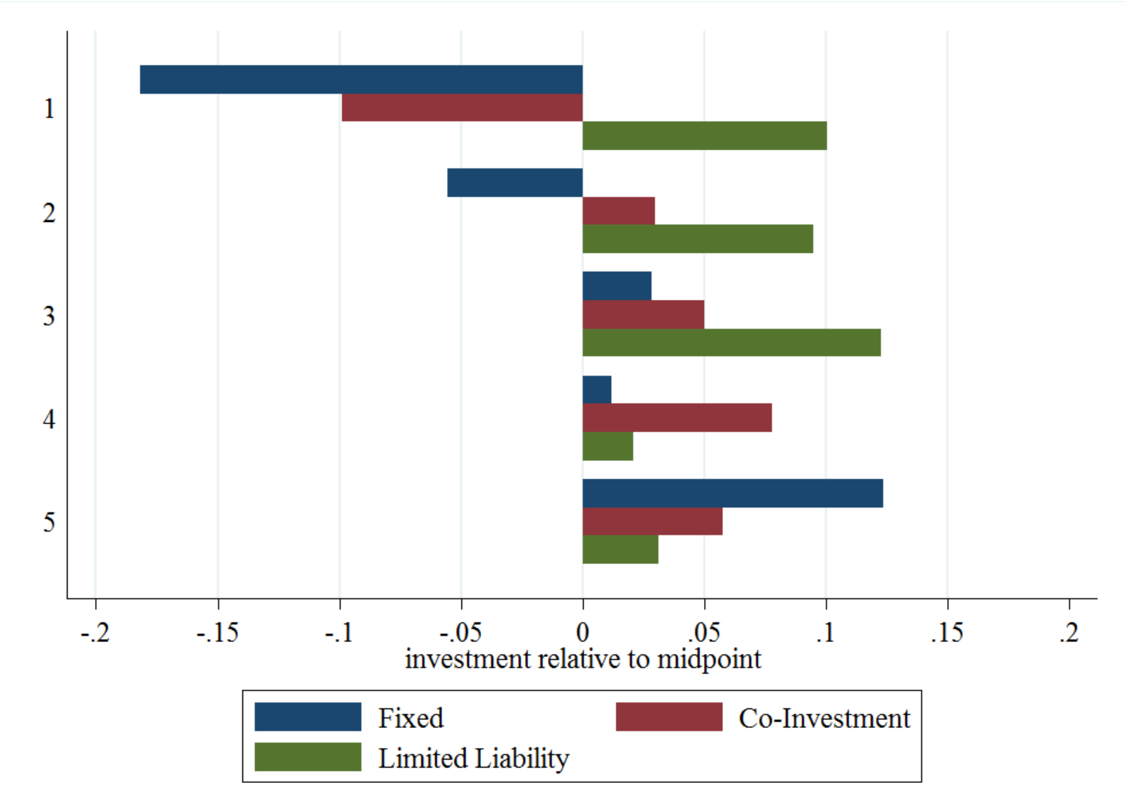
Advisors' Discretion

Despite the fact that advisors in our experiment strongly tailor investments to clients' preferences, they might still react to incentives in a less obvious way. Recall that advisors only learn about the preferred investment profile of their clients. The profiles cover a range of admissible investment levels. Advisors can follow their clients' requests and still use their discretion to their own monetary advantage by choosing investments at the upper end of the requested investment intervals. In the Co-Investment and Limited Liability treatments, this behavior would allow them to both cater to their clients' requests and maximize their own earnings potential.

To analyze whether this behavior occurs in our experiment, we first determine the midpoint of the investment interval that was requested by the client, taking the advisors perception of the investment terms as a basis. We do this for each of the advisors who made an investment decision that is compatible with their client's request. Then we compare the advisors' actual investments to the midpoint of these intervals. Figure 2.7 shows the results for each compensation treatment and for each of the five investment profiles. A value of zero corresponds to the midpoint of the interval, while values of -0.5 and 0.5 would correspond to the lower and upper boundaries of the requested interval.

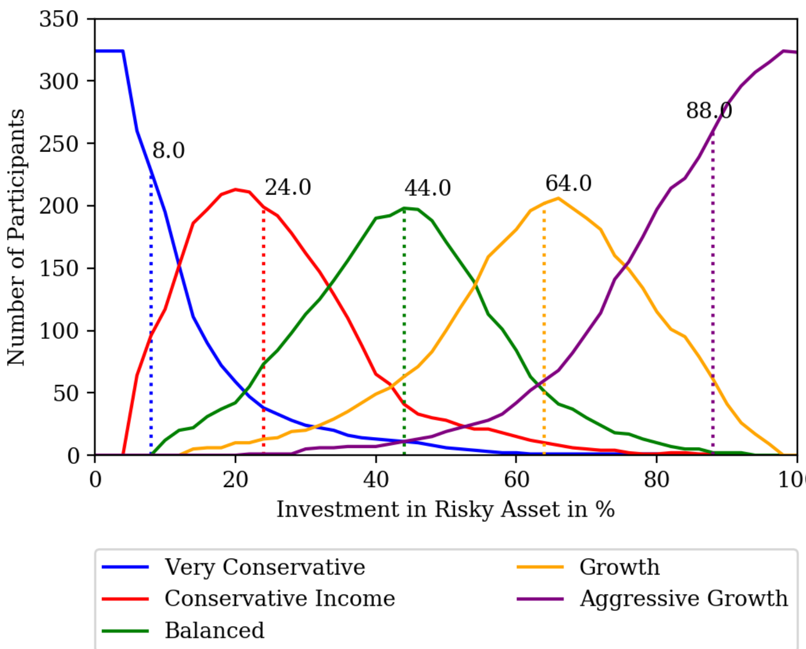
There are visible differences in how advisors use their discretion between the three treatment conditions. In the Fixed treatment advisors seem to use their discretion to conform to the clients' requests as much as possible. For conservative requests they tend to make investments closer to the lower boundary of the interval, while for more risky requests they go beyond the midpoint of the requested interval. In the Co-Invesment treatment we observe a slight shift to the right, with only one of the five requests leading in average to investments below the midpoint of the requested interval. The Limited Liability treatment finally reveals that advisors invest in the upper half of the requested interval in for all of the five possible investment requests. While the effect is strongest for very conservative requests, it is somewhat smaller for investment profiles which imply a higher risk appetite. Clearly, advisors in our experiment react to their own financial incentives, yet they are bound by the moral obligation to their clients.

Figure 2.7: Advisors' Discretion



Notes: The graph shows advisors' investment relative to the requested investment profile for advisors who invested in line with their client's preference. 1 to 5 denote the investment profiles from "very conservative" to "aggressive growth".

Figure 2.8: Perception of Investment Profiles



Notes: For each possible investment share in the risky asset, the graph shows the number of participants who mapped the respective investment profile to the investment share. The individual distributions are labeled with their medians.

2.4.2 Clients' Perspective

The question of how people perceive risks has attracted much research effort. Diacon (2004) compares the perceptions of individual consumers and expert financial advisors and finds strong differences in the perception of financial risks between both groups. It has also been demonstrated that perceptions do not only vary between experts and laymen for financial risks but also for physical or engineering risks (Slovic 1987). Note, that in our experiments, all participants provide their perceptions before they even know that they will take on different roles later on. Combined with our rather homogeneous standard student sample and random treatment assignment, we can only observe heterogeneity in the perception of risk profiles but cannot study systematic differences between advisor and client roles. Figure 2.8 shows the distributions of perceptions of the different investment profiles in our sample. The figure highlights a sizeable overlap of the profiles. For instance, investments in the risky asset between 30% and 60% of the endowment are perceived to match any of the available investment profiles by some participants. Consequently, there is a high degree of heterogeneity in the perception of the different investment profiles and it is far from obvious what they mean to people subjectively. Moreover, the left-shifted medians in Figure 2.10 provide slight evidence for risk aversion in the perception of risky investments.[13] Taken together, the investment profiles commonly used in financial advice appear to be very noisy in their perception.

The most interesting aspect for clients, naturally, is whether they end up with an investment level that is compatible with the preference they indicated to their agents, i.e. whether clients "get what they want". Across all treatments, this is only the case for 43.8% of all clients. Table 2.4 breaks this down by treatment conditions. Each cell shows the percentage of clients that get what they want. For the group treatments, clients seem to get what they prefer more often compared to the single treatments, however none of the differences are statistically significant.

[13] In comparison to a uniform distribution in which each one of the five categories covers 20% of the scale.

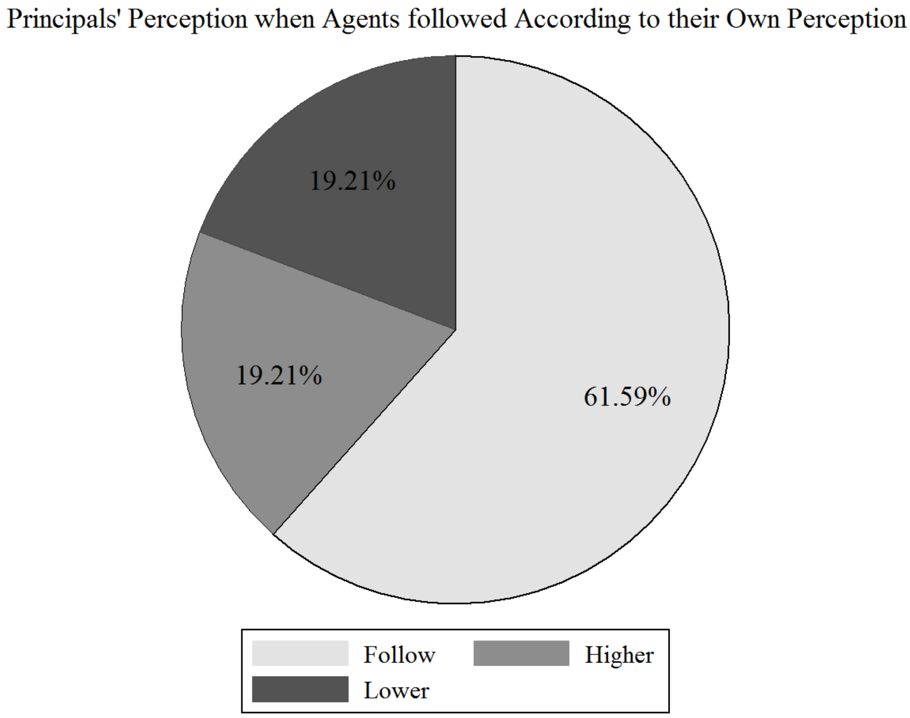
Table 2.4: Share of Clients Who Get What They Want

	Compensation		
	Fixed	Co-Investment	Limited Liability
Single	42.6%	42.6%	40.7%
Group	54.4%	45.6%	45.9%

Notes: The table reports the share of clients who get what they prefer according to their own perception of the investment strategies.

Given this dire picture, one might reasonably ask whether the situation remains the same if we restrict the analysis to those agents, who, according to their knowledge, did their best to implement the profiles preferred by their clients. For only 61.6% of these investments, clients perceive the investment as being compatible with their preferred investment profile. Expressed differently, in 38.4% of the decisions in which agents try to implement their clients' preferred investment profiles, they fail to comply from their clients' point of view. With 19.2% of clients perceiving the decision being lower than preferred and 19.2% perceiving the decision as being higher than preferred there seems to be no systematic deviation, but simply a mismatch in communication on how the strategies are translated into investments in the risky asset. These results are depicted in Figure 2.9.

Figure 2.9: Translation Error



Notes: The graph shows how clients perceive the decisions in which advisors followed their preferred investment profile according to the agents' perception of the profile.

2.5 Discussion

Our results suggest that advisors are in general willing to follow their clients' preferences. Even under unambiguous monetary incentives to take larger risks, advisors strongly consider their clients' preferred investment profiles. Yet, our results hint at an explanation why financial advisors could be perceived to deviate from their clients' preferences. We find evidence for a fundamental problem of communication in financial advice when relying on the use of investment profile terminology. There is a large degree of heterogeneity in the perception of these profiles, which opens up the door for unintended mismatches between advisors' decisions and clients' preferences. The question arises whether this translation error can be reduced by better defining the investment profiles and fostering a common understanding between advisors and their clients.

To shed some light on this issue, we conducted the *Certainty* treatment, which does not leave any room for translation error by design. Disregarding advisors intentions, we find that 42.6% of the clients in this treatment get what they prefer. This is not significantly different from the 40.7% (test of proportions, p = 0.85) in the most comparable Single/Limited Liability treatment. In terms of the outcome clients end up with, the Certainty treatment does not seem to make a difference. However, there are opposing effects acting behind the scenes. In our main treatment, clients may end up with an investment that is compatible with their preferences, despite the fact that their advisor did not intent to implement it. This can happen by chance because of the different perceptions of the investment profiles.

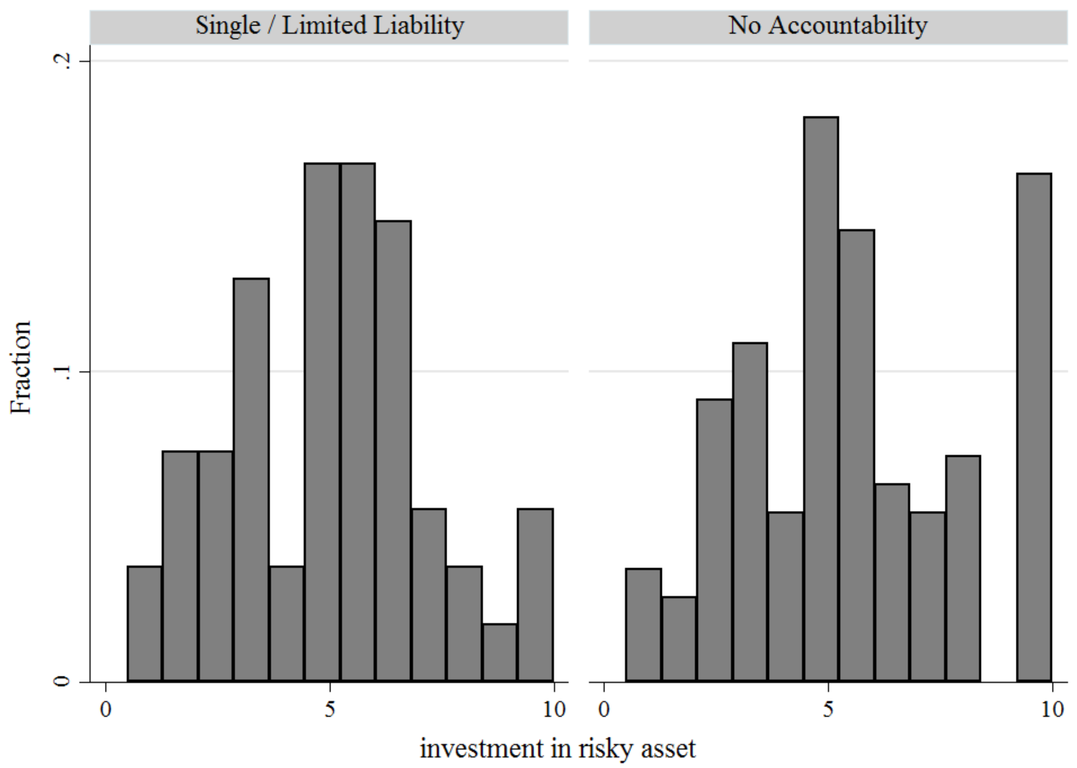
Therefore, a more adequate test of the effects of Certainty is to consider only those observations from our Single/Limited Liability treatment, in which the advisors' intent is to implement the clients' preferences. Limiting the analysis to these observations reveals that the possibility of translation error leads to clients getting an investment they are comfortable with in 46.2% of the cases, substantially less than the 100% in the Certainty case where advisors always correctly implement if it is their intention to do so (test of proportions, $p < 0.01$). However, the Certainty treatment also shows that the absence of uncertainty about the clients' perception of the investment profiles increases the effect of incentives on agents' behavior. Investments in Certainty are higher than in the main Single/Limited Liability treatment after controlling for advisors' and clients' preferences.[14] The share of advisors who invest more than preferred by their clients is significantly larger than the share of advisors who invest less than preferred in the Certainty treatment (test of proportions, 0.44 vs. 0.13, $p < 0.01$). This is not the case for the Single/Limited Liability treatment under uncertainty (test of proportions, 0.30 vs. 0.22, $p = 0.38$).

The consistently high degree to which advisors follow their clients' preferences in our experiment is quite remarkable, yet in line with observations by Ifcher and Zarghamee (2018) and Rud et al. (2019). While observing larger heterogeneity in preferences among clients (Group treatments) appears to increase differentiation as well as investment levels slightly, different incentive schemes do not have much of an effect on investment levels. We hypothesize that the accountability aspect, which is common to all of our main treatments, could be the driving force behind this result. Recall that in all treatment conditions, accountability can stem from multiple sources: First, clients tell advisors how to invest for them. Second, clients can always hold their advisors directly accountable for their decision by sending messages of satisfaction or dissatisfaction with the investment decisions after the fact. Finally, the clear and consistent framing of the

[14] We regress investment in the risky asset on a Certainty treatment indicator and advisors' and clients' preferred investment profiles. The OLS coefficient estimate for the Certainty indicator is 1.19, $p < 0.05$.

experiment as a situation of financial decision-making might instill a heightened feeling of responsibility in agents for their clients' well-being. After all, financial decisions are often considered a matter of mutual trust. To investigate to which degree accountability affects our findings, we conduct our second, additional control treatment *No Accountability*. As described in the design section, we remove all elements which could reasonably make advisors feel accountable for their actions, yet, we do not find a significant increase in the risky investment shares (Kolmogorov-Smirnov test for the equality of distributions: $p = 0.87$. Figure 2.10). It seems that advisors have a feeling of responsibility for their clients, even in the absence of accountability-enhancing design aspects.

Figure 2.10: Investments without Accountability



Notes: The graph shows the distributions of investments in the risky asset for the No Accountability treatment and the most comparable Single / Limited Liability treatment.

Foerster et al. (2017) report that advisor characteristics have a strong influence on portfolio allocations for clients. In fact, advisor characteristics appear to be even more powerful in shaping portfolios than clients' preferences. While both effects persist in our highly controlled laboratory experiments, their strengths change. We find decisions for clients to be predominantly driven by client preferences and estimate advisors' influences to be much weaker. One reason for this difference could be selection. Some financial institutions have been found to select their employees based on behavioral criteria associated with misconduct (Egan et al. 2019). If clients select advisors based on advisor characteristics, or advisors select their target group based on potential clients' characteristics, the strong effects observed by Foerster et al. (2017) can be expected to be dampened in a setting which does not allow for selection in either direction.

2.6 Conclusion

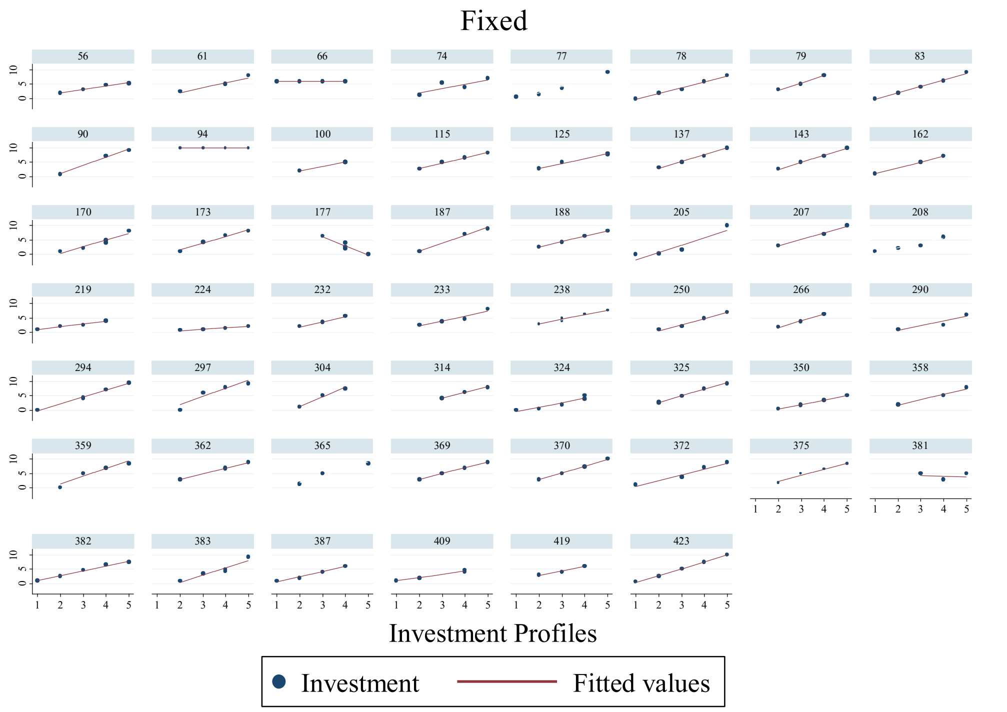
We study whether and how financial advisors shape their clients' investment portfolios in a highly controlled laboratory environment. In general, we observe a high willingness of advisors to follow their clients' preferred investment profiles. Even in light of unambiguous monetary incentives to disregard their clients' preferences, advisors still differentiate between various investment profiles. Yet, clients' portfolios are also affected by their advisors' personal preferences. While our results are qualitatively in line with the findings of Foerster et al. (2017), we do not find the advisors' effects on clients' portfolios to be as pronounced as suggested by their analysis of the empirical data.

By means of our experimental design, we also study the financial advice relationship from another perspective: We examine how clients perceive the investment decisions taken by advisors on their behalf. This reveals that even though financial agents are highly keen to follow their clients' preferred investment profiles, they often fail to achieve their goal from their principals' perspective. One reason for this is that the investment profile terminology, which is often used in financial advice, is very noisy in their perception and people associate them with highly heterogeneous investments into risky assets.

Our results have practical implications for financial advice: In spite of the common perception that financial advisors deviate from their clients' interests, we find advisors to be in general willing to follow their clients' preferences. This still holds under compensation schemes which provide strong financial incentives for advisors to take large risks. However, our findings also point to a fundamental problem in the communication of investment preferences in financial advice. Misunderstanding between advisors and clients are abundant and thus might strengthen the common perception that financial decisions taken by advisors deviate from their clients' interests.

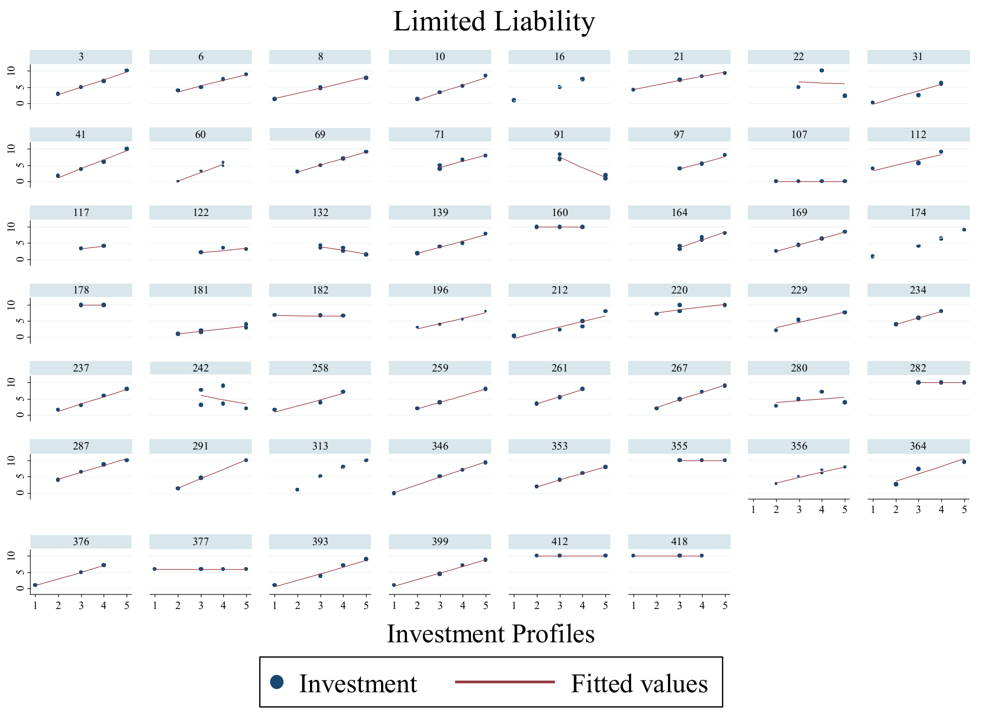
Appendix 2A

Figure 2A.1: Risky Investment by Agent in Fixed treatment



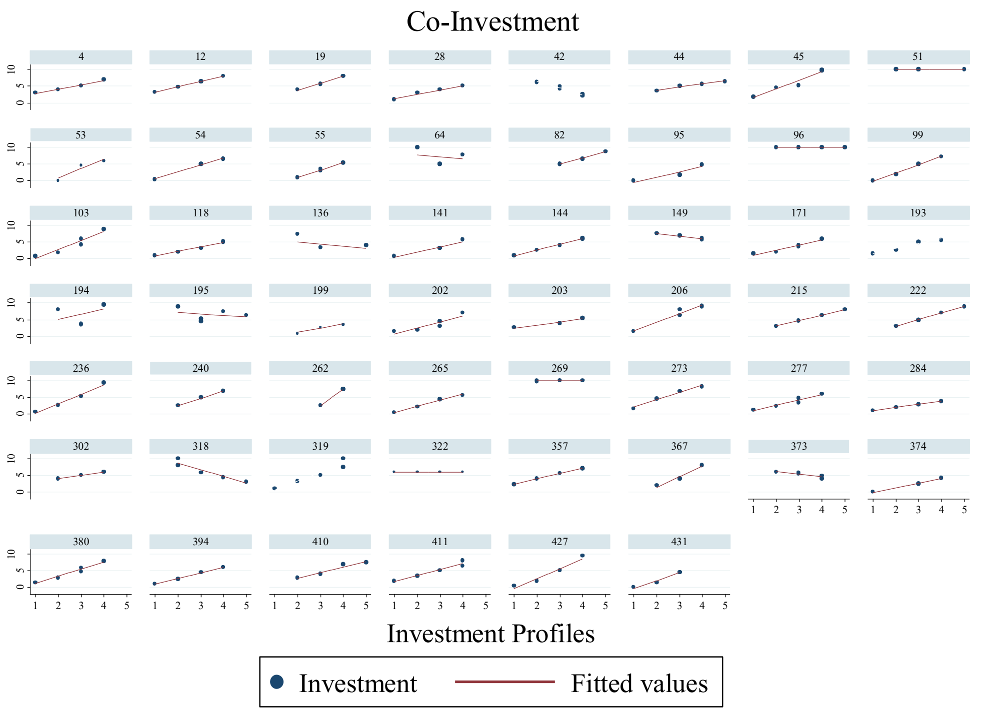
Notes: The graph shows for each participant in the Fixed/Group treatment the investment given the communicated profiles (1 = very conservative, 5 = aggressive growth) of their clients as well as the fitted values.

Figure 2A.2: Risky investment by Agent in the limited liability treatment



Notes: The graph shows for each participant in the Limited Liability/Group treatment the investment given the communicated profiles (1 = very conservative, 5 = aggressive growth) of their clients as well as the fitted values.

Figure 2A.3: Risky Investment by Agent in the Co-investment Treatment



Notes: The graph shows for each participant in the Co-Investment/Group treatment the investment given the communicated profiles (1 = very conservative, 5 = aggressive growth) of their clients as well as the fitted values.

Chapter 3

Good Decision vs. Good Results: Outcome Bias in the Evaluation of Financial Agents

Abstract. We document outcome bias in situations where an agent makes risky financial decisions for a principal. In three experiments, we show that the principal's evaluations and financial rewards for the agent are strongly affected by the random outcome of the risky investment. This happens despite her exact knowledge of the investment strategy, which can therefore be assessed independently of the outcome. The principal thus judges the same decision by the agent differently, depending on factors that the agent has no influence on. The effect of outcomes persists in a setting where principals communicate a preferred investment level. Principals are more satisfied with the agent after a random success when the agent did not follow the requested investment level, than after a failed investment that followed their explicit request.[15]

[15] This chapter is co-authored by Monique Pollmann, Jan Potters and Stefan T. Trautmann.

3.1 Introduction

Whenever the quality of a decision is evaluated after its consequences have played out and have become public knowledge, there is a chance of falling prey to outcome bias. *Outcome bias* describes the phenomenon by which evaluators tend to take information about the outcome into account when evaluating the quality of a decision itself (Baron and Hershey 1988). This tendency is problematic for two reasons. First, the evaluator has available a different information set than the decision maker, who typically faces uncertainty at the time of her decision. Second, a good outcome might derive from a bad decision, and a bad outcome might derive from a good decision.[16] Evaluation of outcomes may therefore be questionable and may lead to suboptimal future decisions if decision makers follow strategies that were successful only by chance (e.g., Bertrand and Mullainathan 2001, for managerial performance; or Sirri and Tufano 1998, for investors' mutual fund choices).[17]

The consideration of potentially irrelevant outcome information in the evaluation of decision quality has been documented in a wide variety of settings including medical advice, military combat decisions and salesperson performance evaluation (Baron and Hershey 1988, Lipshitz 1989, Marshall and Mowen 1993). In these early studies, participants were asked to evaluate the quality of a decision described in hypothetical scenarios differing in featuring a favorable, an unfavorable, or no outcome at all. Later studies on peer review of scientific publications and strategies in professional football move away from scenarios and towards actual decisions as the basis for evaluation (Emerson et al. 2010, Lefgren, et al. 2015). Relatedly, there is a strand of literature on allocator-responder games with a 'trembling hand' condition, in which responders can infer allocators' intentions, but actual allocation outcomes may deviate from intentions by chance. Cushman et al. (2009) find that responders hold allocators accountable for unintentional negative outcomes, but knowledge of their agents' intentions moderates the

[16] Consider for example a decision between a safe payment and a prospect with positive expected value larger than the safe option, but of substantial variance. A decision maker instructed to make risk-neutral decisions should choose the risky prospect over the safe option. Yet the outcome might turn out unfavorable and lower than the safe option. A negative evaluation on the basis of the bad outcome seems unwarranted.

[17] It is important to recognize that outcome effects do not always constitute biases. The literature originating from Baron and Hershey (1988) typically speaks of an outcome bias only if responsibility for the outcome is inappropriately assigned to decision makers. We follow this interpretation.

effects. These findings are supported and augmented by further studies, e.g. by Murata et al. (2015) and Sezer et al. (2016).

Investors choosing investment funds that have been successful in the past (mostly by chance) is an important policy issue in finance. However, in the field of financial economics, there is little controlled experimental evidence yet on whether such behavior is potentially related to outcome bias: field data typically cannot separate outcome bias from other effects. Our paper is the first to study outcome bias in such a financial investment context, using controlled lab experiments. We focus on client-advisor relationships typical to investment settings (framed in terms of a principal-agent situation), and show that outcome bias is prevalent in such settings and leads to biased assessment of the quality of the advisor's investments by the client. To the best of our knowledge, we are also the first who identify positivity bias in the context of outcome bias. Such positivity bias is consistent with the above described field evidence on investors following mutual funds successful in the past.

Why do we need to study outcome bias in financial investment settings if there is already broader evidence on the bias? For behavioral experiments, seemingly small changes to decision situations can have pronounced behavioral consequences. Especially concerning cognitive biases, transferability from one situation to another, even if they appear to be highly similar, cannot be taken for granted (Crusius et al. 2012). For example, Charness et al. (2010) show that the introduction of mild incentives significantly reduces violations of the conjunction principle compared to an otherwise identical, but unincentivized decision situation. In addition, Lefebvre et al. (2011) highlight that the ratio bias phenomenon is sensitive to changes in the decision-making environment as well as the incentive structure. The current paper concerns the robustness of the outcome bias phenomenon. We assess the prevalence and implications of the outcome bias in financial decisions with agency, employing a variety of different incentive conditions and assessment methods by the evaluator.

In three experiments, we document evidence on outcome bias in the evaluation of financial agents who take investment decisions for another person. In Experiment 1, the principals' assessments of the agents' decisions have direct monetary consequences for principals and agents, and potentially affect agents' future decisions. We compare a situation where principals can observe both the decision itself and the resulting outcome, to a situation where only the investment decision is known but no outcome information is available yet.[18] We observe that a tendency toward ex-post outcome-based evaluations exists even in situations where (1) the principal has a clear financial incentive to reward good decisions, not lucky good outcomes; and (2) where there is perfect information about the decision and the situation in which it was made.

To control for potential design-specific social-preference effects that reduce the generalizability of our results, we probe the effect of outcome-based evaluations of known processes in Experiment 2. We find that even in the absence of potential social-preference effects, principals' judgments of agents' observable investment decisions are strongly affected by the random outcome on which the agent has no influence. In particular, principals become satisfied with investment *decisions* after positive *outcomes* even if they initially strongly disliked the decision (in the absence of the outcome information, i.e., before the uncertainty is resolved). This positivity bias is consistent with findings by Casal et al. (2019), but was unexpected given the previous demonstrations of the predominance of negative outcome effects (Gurdal et al. 2013, Ratner and Herbst 2005). The current findings suggest that financial agents seem to benefit from the rule that the *result justifies the deeds*. Inspired by our results in Experiment 1, Oliveira et al. (2017) also study the role of wealth differences for outcome effects (see our analysis in Section 3.2.4). Consistent with our Experiment 2 results, the authors find no support for a wealth-based explanation of outcome effects.

In Experiment 3, we replace the principals' implicit assessment of the agents' decision strategies by the principal's explicit demand of a certain investment level. After observing the invested amounts (which are indicative of the agents following or not following their clients' requests) and the investment outcomes, principals send messages conveying their satisfaction with the investment decisions to their agents. This setup enhances the situation modeled in Experiment 1 by providing (1) the agents with information about what their clients consider a 'good decision' and thus (2) the principals with an obvious benchmark for evaluating the agents' decisions. We observe that satisfaction with the decision is increased by the agent conforming to the principal's wish, but find an even larger effect of the random outcome on satisfaction with the investment.

[18] The former condition is similar to experiments in Gurdal et al. (2013) where players were rewarded for choosing a risky or a safe lottery for another player. Counterfactual outcomes were available to judges and had an influence on rewards. Below we discuss Gurdal et al.'s interpretation in terms of blame in the light of our results.

In sum, we demonstrate that outcome bias is present in financial decisions by agents, in which the evaluator is directly monetarily affected by both the decision and its evaluation in terms of monetary rewards paid (Experiment 1). We show that the phenomenon persists in situations where wealth effects and social preference considerations cannot play a role, and separation of outcome and decision process evaluation is strongly emphasized (Experiment 2). Finally, we highlight that even if principals communicate explicit investment-level demands, they still fall prey to outcome effects if the agents do not follow their demands (Experiment 3). Recognizing, that past experience can bias future evaluations (cf. *rater bias* in Müller and Weinschenk, 2015), such systematically biased assessments of the quality of agents' decisions are clearly undesirable.

The remainder of the paper is laid out as follows. Sections 3.2 to 3.4 describe the methods and the results of the three experimental studies. Each section also includes a short discussion of the respective results. Section 3.5 concludes the paper with a general discussion of the role of outcome bias in financial agency.

3.2 Experiment 1

3.2.1 Methods

We use data from Pollmann et al.'s (2014) experiment on risk taking by agents under accountability.[19] That paper investigates how behavior of financial agents differs between situations in which the principals either reward their agents solely based on invested amounts or on invested amounts and outcome information. Pollmann et al. (2014) is not concerned with the behavior of principals, but, by necessity, implements treatments which are suitable to study their behavior as well.

The Gneezy and Potters (1997) investment task is used in the experiment, in which decision makers are asked to divide an initial endowment of 100 points between a safe and a risky asset. The safe asset has a return of 0%. In contrast, the risky asset has a return of +250% with a probability of 1/3 and a return of -100% with a probability of 2/3, creating a prospect with a positive expected return of +16.67%.

[19] The data and instructions of all three experiments are part of the supplementary material available online at https://ckgk.de/files/thesis/ch3_outcome_bias.zip.

There are two types of players matched in pairs of two: a principal who is the owner of a 100-point endowment; and an agent, whose task it is to invest the principal's endowment using the above-described technology. The investment portfolio set up by the agent is fully owned by the principal. Both players receive an additional fixed payment of 100 points each. After the agent made her investment decision, the principal is given the opportunity to reward the agent by transferring between 0 and 100 points from this additional payment to the agent. This ensures that principals can give any reward, independently of how their payoffs from the agents' investment decisions turn out. Points not transferred remain with the principal. The agent receives this reward in addition to her fixed payment of 100 points. Employing a between-subject design, we compare two treatments that differ in terms of the information the principal has available when she is given the opportunity to reward the agent for her decision.[20] When making her decision of how many points to transfer as a reward in treatment REWARD BEFORE, the principal knows the agent's investment decision (number of points invested in risky and safe), but not the realized return of the risky asset. In treatment REWARD AFTER, both the agent's investment decision and the outcome of the risky prospect are communicated to the principal before she has the opportunity to reward the agent.

The described tasks (investment – reward) are statically repeated five times with fixed principal-agent pairs. This setup increases the importance for the principal to reward investment decisions that is in line with her preferences (not outcomes that are positive), because the same agent will make another investment decision after the reward is given. At the end of each round, payoffs for each player are transferred to her experiment account and cannot be used in the experiment anymore. New endowment and investment funds are provided for each round, ensuring that although wealth is accumulated over time, the decision set remains identical.

3.2.2 Setting and Summary Statistics

The experiment was programmed in z-Tree (Fischbacher 2007) and conducted at CentErlab, Tilburg (NLD). Roles were assigned randomly, partner identities were kept secret, and decisions were made anonymously with no communication between

[20] The experiment consisted of two more treatments where participants made investment decisions for their own account, and where they made decisions for others without the possibility of reward. These are discussed in Pollmann et al. (2014).

principals and agents. Participants received instructions in writing as well as on screen and had to complete a set of mandatory comprehension questions. The sessions began only after every participant had correctly answered these questions. The research question was not revealed to participants at any time. Points were exchanged for 0.01€ each at the end of the experiment.[21]

A total of 134 students participated in the part of the experiment relevant for this paper (34 principal-agent pairs in treatment REWARD BEFORE and 33 principal-agent pairs in treatment REWARD AFTER). At the time of the experiment, participants were on average 22.5 years old, 37% of them were female and 36% of Dutch nationality. We asked participants about their major field of studies, which revealed 55% economics, 37% business, and 2% psychology students in our sample. Table 3B.1 in Appendix 3B presents the summary statistics in detail.

3.2.3 Results

Comparing the rewarding behavior of principals in treatment REWARD AFTER in situations in which the risky asset yielded a positive random outcome to situations in which it yielded a negative one, we observe substantial outcome effects. Pooling observations from all rounds, we find average rewards of 28.78 (SD = 4.36) when favorable outcomes are observed, versus 10.54 (SD = 1.82) when unfavorable rewards are observed.[22] As a placebo test, we make the same comparison for treatment REWARD BEFORE. Here we find average rewards of 18.72 (SD = 3.12) for favorable random outcomes, versus 18.94 (SD = 2.47) for unfavorable ones.[23]

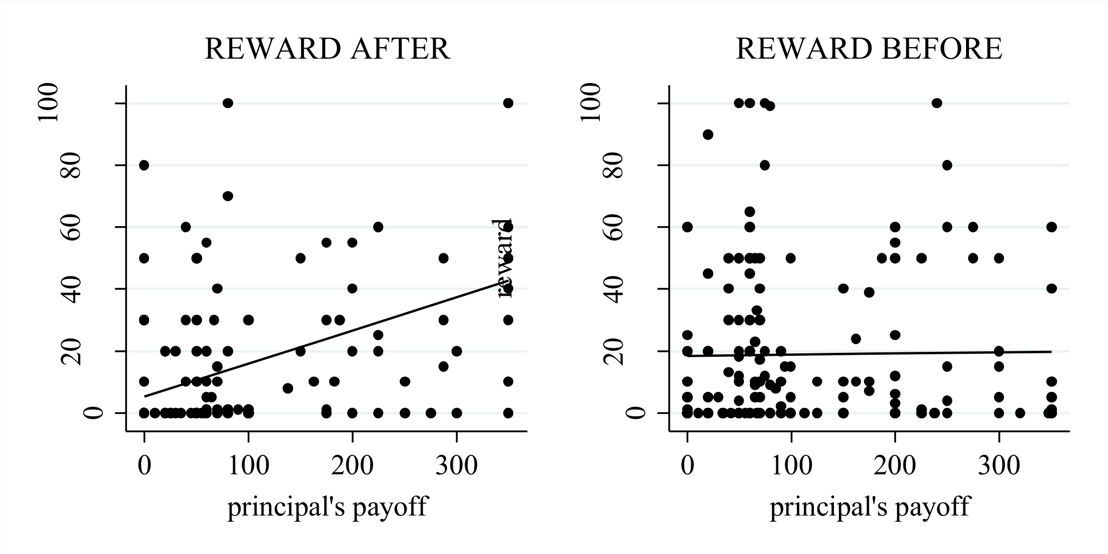
For principals who received information about the investment decision and outcomes, we furthermore see a significantly positive correlation between their own payoff and the reward they pay to their agent (Figure 3.1, left panel, $\delta = 0.45$, $p < .001$). We do not find a positive correlation if the principal had to reward the agent before knowing the outcomes of the risky investment (Figure 3.1, right panel, $\delta = 0.02$, $p = .83$).

[21] Participants could earn an additional 100 points in a belief elicitation task, which is not discussed in this paper. On average, participants earned 7.93€. We provide the complete instructions to this experiment as part of the supplementary material.

[22] Two-sample, two-sided t-test, $t(163) = -4.56$, $p < .001$, $d = -0.76$. When rewards are compared separately for each period, the difference is significant in three out of five periods. We account for the repeated structure in the multivariate analysis below.

[23] Two-sample, two-sided t-test, $t(168) = 0.06$, $p = .956$, $d = 0.01$. The difference in average rewards is neither significant when observations are pooled nor when periods are treated separately.

Figure 3.1: Relation between Principals' Payoffs and Rewards for the Agents



Notes: All periods included; scattered observations with linearly fitted line.

Table 3.1: Relation between Principals' Payoffs and Rewards for the Agents

	AFTER Size of Reward	AFTER Size of Reward	BEFORE Size of Reward	BEFORE Size of Reward
Principal's Payoff	0.0906 *** (0.0134)	0.0903 *** (0.0129)	-0.0011 (0.0117)	-0.0015 (0.0117)
Controls	No	Yes	No	Yes
# principals	33	33	34	34
# observations	165	165	170	170

Notes: Random effects Tobit regression; average marginal effects reported; standard errors in parentheses; *** denotes significance at the 0.1% level; controls are: age, gender, field of study and Dutch nationality.

In order to estimate the size of the effect as well as to control for repeated observations and personal characteristics of the participants, we probe these findings in a multivariate analysis. For each treatment, we employ a separate Tobit panel regression to regress the size of the reward on the principal's payoff and a constant. In a second step, we test the robustness of the results by including controls for age, gender, Dutch nationality,[24] and the field of study. As coefficients are hard to interpret in non-linear models, we report the more convenient average marginal effects in Table 3.1. The regression analyses confirm that absent information on realized outcomes, there is no effect of the principals' payoffs on rewards. However, once outcomes are available, there

[24] The experiment was run in the Netherlands, but with a significant group of foreign students. Since rewarding behavior may vary across different cultural backgrounds we control for Dutch versus foreign students here.

is a significantly positive effect of payoffs on rewards: on average, a unit increase in payoff leads to an increase of 0.09 points in reward.[25]

Because high payoffs obtain from favorable random draws for the risky investment, we next test whether it is the observation of a success or failure per se that drives the above effect, or whether the effect runs mainly through the size of the outcome. We thus repeat the above analyses, now including as covariates the amount invested in the risky asset, an indicator for a favorable outcome (investment success) and the interaction of these variables. Results are shown in Table 3.2. If both the investment decision and the outcome are observable (REWARD AFTER, Table 3.2 upper panel), we can report two results: First, the reward in the case of observing a favorable outcome is on average 16.60 points higher than in the case of observing an unfavorable outcome. Second, if the outcome is favorable, the effect of the amount invested on the reward is positive and highly significant. A unit increase in risky investment leads to an average increase in reward of 0.46 points. If the outcome is unfavorable, the effect of the amount invested in the risky asset on the reward is not significantly different from zero. That is, rewards are driven by success in REWARD AFTER, and only in the case of success does the amount invested, and therefore the actual payoff to the principal, affect the size of the reward. In the case of a failure, the correlation between principal's payoff (which then depends inversely on the agent's investment) and the reward is close to zero and non-significant.

If only the amount investment is observable by the principal (REWARD BEFORE, Table 3.2 lower panel), we do not find a statistically significant effect of the invested amount on rewards. The placebo test of the effect of the favorable outcome is also insignificant. As shown in Table 3.2, all results are robust to the inclusion of demographic controls.

[25] The marginal effect of an increase in payoff on reward is significantly different from zero at all levels of payoff and monotonically increasing from 0.06 to 0.15. Graphs of the marginal effects are available from the authors upon request. All results are robust to using a linear panel OLS regression with standard errors clustered on the individual level instead of the non-linear tobit model.

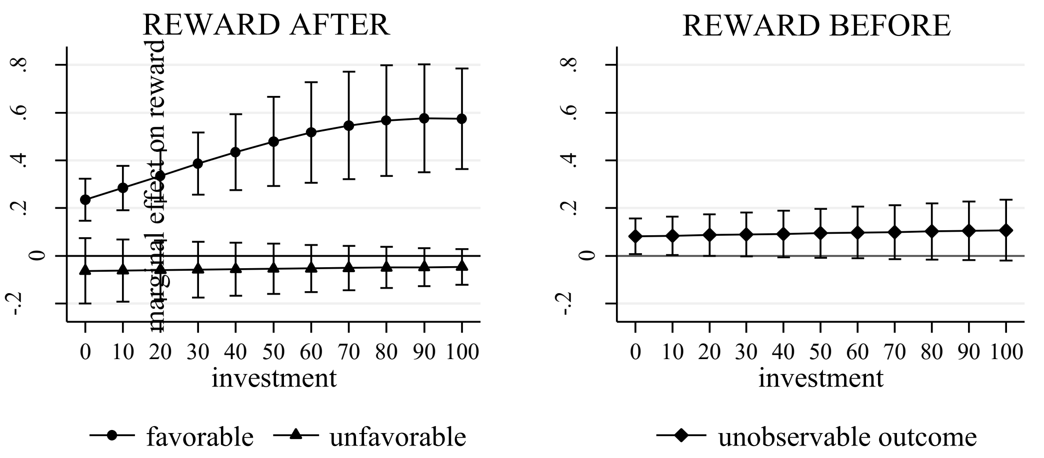
Table 3.2: Relation between Agents' Risky Investment and Rewards

REWARD AFTER	Size of Reward	Size of Reward
Investment if observing…		
…a favorable outcome	0.4602 *** (0.0785)	0.4845 *** (0.0798)
…an unfavorable outcome	-0.0531 (0.0506)	-0.0457 (0.0494)
Favorable Outcome	16.5962 *** (2.6101)	16.1321 *** (2.5799)
Controls	No	Yes
# principals	33	33
# observations	165	165
REWARD BEFORE	Size of Reward	Size of Reward
Investment	0.0943 (0.0522)	0.0985 (0.0519)
Favorable Outcome	-0.5752 (2.4931)	-0.6896 (2.4094)
Controls	No	Yes
# principals	34	34
# observations	170	170

Notes: Random effects Tobit regression; average marginal effects reported; standard errors in parentheses;*** denotes significance at the 0.1% level; controls are: age, gender, field of study and Dutch nationality.

Figure 3.2 provides further insights by plotting the marginal effects on rewards for each investment level separately. For REWARD AFTER, it shows that in the case of an unfavorable outcome there is indeed an insignificant effect at all investment levels; in contrast, for a favorable outcome the marginal effect is increasing in the investment level. Moreover, the figure confirms the signs of the Tobit interaction terms in Table 3.2. For the case of REWARD BEFORE, the figure shows that investment has a significantly positive effect only at very low investment levels.

Figure 3.2: Marginal Effects of Agents' Risky Investment on Reward



Notably, the observed pattern of rewards is not consistent with a general wealth effect. If more wealth, i.e. a higher payoff to the principal, generally translated into higher rewards for the agents, we would observe a negative effect of the size of the investment on rewards for unfavorable investment outcomes: wealth is decreasing in the investment in this case. The observed pattern is also robust if we restrict the analysis to situations in which principals clearly take the agents' payoffs into consideration by paying non-zero rewards. In particular, the marginal effects of investment level on rewards are close to zero for unfavorable investments, and positive and increasing after an investment success. Taken together, only a wealth effect conditional on observing a positive outcome is consistent with our results. This conditioning is exactly what outcome bias implies. A similar argument applies to the possibility that instead of wealth, it is an implicit experimenter demand that drives the outcome bias: available information should be used in the determination of the reward. If this were the case, as we have shown, only positive information about a successful outcome would create experimenter demand; principals are not responsive to experimenter demand caused by increasingly negative outcomes after an unsuccessful investment. That is, this model would imply an outcome bias in participants' assessment of experimenter demand.

3.2.4 Outcome Bias and Social Comparison

We observe that principals strongly base their rewards on observed outcomes when these are available. In particular, principals reward favorable chance outcomes, and additionally reward higher investments conditional on hindsight that larger investments were a good decision. Given that (i) the outcome is not under the control of the agent and (ii) the principal has full information about the agent's decision process (i.e., amount invested in the presence of uncertainty), it seems difficult to justify this focus on outcomes.

Despite our finding that wealth level effects cannot account for the observed pattern of rewards after favorable and unfavorable investment outcomes, social comparison may still loom large in the current experiment, and may add to the observed outcome effect. To gain some insight into this potential channel behind the observed outcome bias, we analyze the data of Experiment 1 within the context of social preference models. We consider the model proposed by Fehr and Schmidt (1999) for the case of observable outcomes (REWARD AFTER), and the model proposed by Trautmann (2009) for the case of unobservable uncertain outcomes (REWARD AFTER). We assume that after a high payoff to the principal, she might be more inclined to give a higher reward to the agent to make payoffs more equal. That is, we assume that the principals are averse to advantageous inequality. In Appendix 3A, we show that the outcome-based model cannot explain the observed patterns of reward in REWARD AFTER for a fixed distribution of inequality aversion parameters. This reflects our above observation of an absent link between investment and reward after an unsuccessful investment. In contrast, for REWARD BEFORE, distributions of inequality aversion parameters can be constructed to fit the observed pattern of rewards.

Despite the failure of inequality aversion models to account for the pattern of rewards when outcomes are observed, feelings of fairness will obviously be important in many situations outside the lab. It can feel inappropriate not to reward a successful manager despite him profiting from random events occurring in the market. Similarly, a blackjack player may tip the dealer more generously after a good hand. Social comparison motives may thus also loom large in the evaluation of agents outside the current experimental setup and may contribute to outcome bias: it may simply feel inappropriate not to reward an agent after a good result, even if the way the result was obtained would otherwise be judged negatively. Conditional on some reward being appropriate, the size of the reward may in fact depend on social comparison considerations (e.g., equality considerations). However, to probe the generalizability of the outcome bias in financial agency settings where social preferences may be less directly relevant, we conducted a second experiment that excludes social preferences and gives further insights into the interaction of outcome and decision-process evaluations.

3.3 Experiment 2

3.3.1 Methods

The second experiment elicits judgments of an agent's investment decision, and the resulting investment outcome, by a principal. We employ an unincentivized vignette format in this experiment for two reasons. First, the design allows us to exogenously manipulate different investment levels (risky vs. safe) and different uncertain outcomes (success vs. failure). Second, by directly eliciting measures of satisfaction we prevent social preference issues that become relevant in the allocation of principals' and agents' payoffs with decision-based monetary payoffs.

In this experiment, we present hypothetical scenarios involving a financial advisor who is tasked to allocate $10,000 between a safe and a risky asset for the participant. The scenario is identical to the Gneezy-Potters task used in Experiment 1.[26] We employ two possible allocations, with either low ($1,500) or high ($8,500) investments in the risky asset and the remainder being invested in the safe asset. In addition to the general scenario and the description of the two assets, we present the agents' investment decision and, depending on the treatment, the outcome of the risky investment. Participants are asked to indicate separately their satisfaction with the investment decision and, if known, the outcome on a 7-point Likert scale[27]: "How satisfied are you with the <u>investment decision</u> the adviser took for you?" and "How satisfied are you with the <u>outcome</u> of the investment decision the adviser took for you?" (emphasis in the original). That is, the survey carefully distinguished between the decision to investment a certain level, and the success or failure of the investment. Without explicitly asking for both aspects separately, participants might have construed the term "decision" in a way that comprises the resulting outcome (Blank et al. 2015). By separating the two aspects, participants can signal discontent with a decision that does not fit their risk appetite, while at the same time acknowledging their happiness about the outcome (or vice versa).

The scenario, the advisor's decision, as well as outcome information are presented on the same screen as the questions regarding participants' satisfaction. The experiment concludes with a short questionnaire collecting age, gender, education level and current occupation. Table 3.3 provides an overview of the six between-subjects conditions as

[26] We familiarize participants with the investment situation by having them calculate the payoffs for different outcomes of a hypothetical $5,000 investment in each type of asset.

[27] Our Likert scales range from „very dissatisfied“ (1) to „very satisfied“ (7). Numbers are not shown.

well as the respective number of observations. In the current experiment, the ‘unknown’ condition corresponds to a situation of REWARD BEFORE, and the ‘positive’ and ‘negative’ conditions correspond to the situation of REWARD AFTER, in Experiment 1.

Table 3.3: Six Treatments in Experiment 2

Outcome / Investment	unknown		positive		negative	
	name	# obs.	name	# obs.	name	# obs.
Low investment	L?	51	L+	50	L–	48
High investment	H?	51	H+	48	H–	49

3.3.2 Setting and Summary Statistics

In total, 297 volunteers, recruited on Amazon Mechanical Turk, completed the online experiment and received a compensation of $0.50 each for their participation, which took a little more than 5 minutes on average. The actual survey was implemented using SoSci Survey (Leiner 2014). As part of the study description on Amazon Mechanical Turk, we mentioned being “interested in how people judge certain situations”[28], but neither revealed the research question nor that there were different conditions. We made sure that participants could only take part in the study exactly once and restricted the sample to participants from the US to avoid language barriers and ensure a minimum of homogeneity in the cultural background.

With an average age of close to 39 years, our online sample is older and more heterogeneous than the student samples participating in the laboratory experiments. At the same time, online participants are also much more diverse in their academic background. Only 3% and 14% are trained in economics and business respectively, while 4% are psychologists. Females comprise 47% of the sample.

3.3.3 Results and Discussion

Outcome satisfaction ratings for the four treatments in which the outcome of the investment decision was available to the participants are shown in Table 3.4 (upper panel) and Figure 3.3 (upper half). As expected, participants indicate significantly higher satisfaction with positive compared to negative outcomes for both low and high investment amounts in the risky asset.

[28] The instructions to Experiment 2 are available in the supplementary material.

Next, we consider participants' satisfaction with the investment *decision* itself, rather than with the random *outcome*. Table 3.4 (lower panel) and Figure 3.3 (lower half) summarize the findings. As a first result we find support for the common observation of risk aversion in the current investment setting with potential losses: mean satisfaction with the decision is generally higher for low investment compared to high investment in the absence of outcome information (two-sample, two-sided t-test, M = 4.39 vs. M= 2.39, t(100) = 6.10, p < .001, d = 1.21). Rating patterns in the unknown outcome treatments further support this observation: For the low investment in the risky asset, the distribution of ratings is almost uniform, while it is clearly skewed towards a negative evaluation in the high investment case (see Figure 3.3).

We now consider decision satisfaction ratings across the different outcomes for each investment level. In the absence of outcome bias, we would expect there to be no differences in decision-process satisfaction ratings. Participants were given information on both the decision and the outcome, and had the possibility to indicate satisfaction separately for outcomes and decisions. However, consistent with outcome bias, we observe significantly higher ratings of the same decision after a randomly obtained good investment outcome compared to a bad investment outcome, for both investment levels.

Table 3.4: Investment Outcome and Decision Satisfaction

Treatment	L?	L+	L–	H?	H+	H–
Outcome	n.a.	6.24	3.38	n.a.	6.58	1.35
		***				***
Decision	4.39	5.58	4.54	2.39	4.85	1.73
	***		***	***		***
					*	

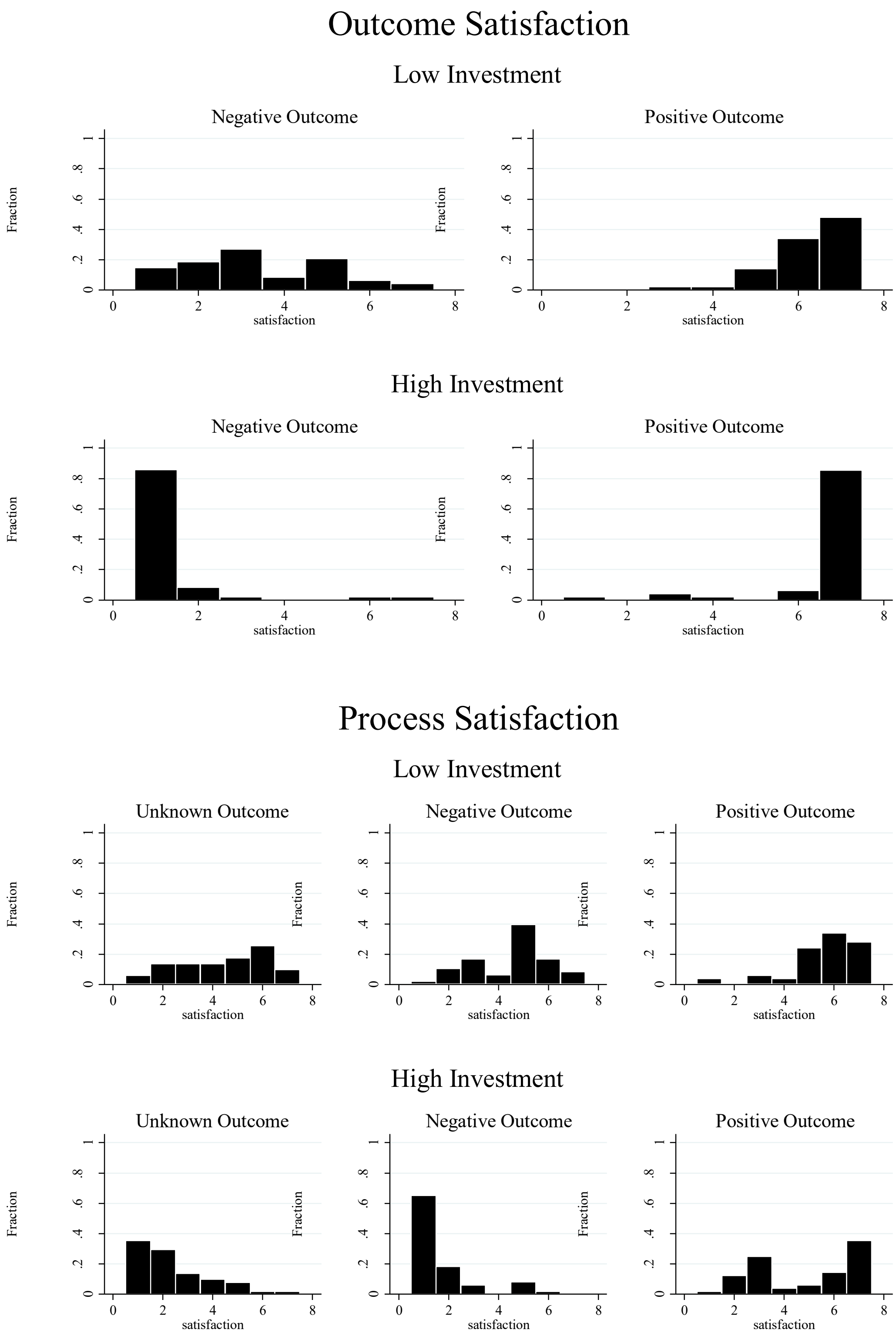
Notes: Means and two-sided t-tests on satisfaction ratings reported; */*** denote significance at the 5%/0.1% level. Individual test statistics below:
L+ vs. L–: t(96) = -10.52, p < .001, d = -2.13.
H+ vs. H–: t(95) = -21.65, p < .001, d = -4.40.
L? vs. L+: t(99) = 3.67, p < .001, d = 0.73.
L+ vs. L–: t(96) = -3.46, p < .001, d = -0.70.
L? vs. L–: t(97) = 0.45, p = .657, d = 0.09.
H? vs. H+: t(97) = 6.87, p < .001, d = 1.38.
H+ vs. H–: t(95) = -8.99, p < .001, d = -1.83.
H? vs. H–: t(98), p < 0.05, d = -0.46.

Comparing the evaluation of the decision process in the presence of outcome information to the situations where participants judged the process in the absence of outcome information, we observe that good outcomes have a strongly positive effect, while negative outcomes have a more modest negative effect on decision-process judgments. These results are confirmed in a multivariate analysis (Appendix 3B). Observed outcomes have an effect both on outcome satisfaction and on decision-process satisfaction. The effect is stronger for outcome satisfaction, but still economically and statistically significant for decision-process judgments. Positive effects for good outcomes on process judgments are more pronounced than the negative effects of bad outcomes, for both investment levels. The latter effect is consistent with rewarding behavior in Experiment 1 that also hints at a positive bias. The absolute difference in average rewards between REWARD BEFORE (unknown outcome) and REWARD AFTER (known outcome) is larger for favorable (9.92) than for unfavorable outcomes (8.32); however, the difference is not significant.[29]

Experiment 2 results also challenge a possible explanation of the effect in terms of experimenter demand. As we observed, outcomes are very salient, both good and bad ones. If participants believe that every piece of information provided by the experimenter is relevant for the situation at hand and should inform their decision, we expect that explicitly mentioning successful or unsuccessful outcomes should have a comparable effect on process evaluations. This is not the case. Positive outcomes clearly have a stronger effect and in the low investment condition, the comparison between unknown outcomes and negative outcomes is even insignificant. That is, if there were experimenter demand, it would be highly asymmetric, and driven by outcomes. Information per se does not seem to affect process evaluations.

[29] REWARD BEFORE, $M = 18.86$ vs. REWARD AFTER, favorable outcome, $M = 28.78$: two-sample, two-sided t-test, $t(222) = -2.35$, $p < .05$, $d = -0.37$. REWARD BEFORE, $M = 18.86$ vs. REWARD AFTER, unfavorable outcome, $M = 10.54$: two-sample, two-sided t-test, $t(279) = 2.96$, $p < .01$, $d = 0.36$. Note that the identification of positive bias is more difficult in Experiment 1 because of the endogenous amount of investment.

Figure 3.3: Investment Outcome and Process (Decision) Satisfaction



Notes: Satisfaction ratings from 1 “very dissatisfied” to 7 “very satisfied”.

In addition, recall that we ask participants to evaluate their satisfaction with the investment decision and the outcome in two separate questions. We do not only make the difference salient and give participants the opportunity to cleanly distinguish between the two aspects, but even specifically demand them to do so. Clearly, participants' outcome satisfaction is driven by the observed outcome. Yet, even if explicitly asked, they are to a large degree unable to prevent outcome information from affecting their decision process evaluation, as soon as outcome information is available.

In sum, we find clear evidence for the outcome bias in the judgment of agents' investment decisions. Investment decisions were fully observable, and social preference effects were excluded by design.

3.4 Experiment 3

3.4.1 Methods

In the first two experiments, principals had to judge how satisfied they were with the investment decision without having to commit to what they consider a good decision ex-ante. Although this is a realistic feature in many applied settings, it might have amplified the outcome focus if people construct their preference ex-post. In Experiment 3, we reduce this uncertainty about the principals' ex-ante preference by letting them state their preferred investment strategy to their agents, who afterwards take a Gneezy-Potters (1997) investment decision for them.[30] Note that the investment decision by the agent may be influenced by her own financial interest, which may deviate from the principal's preferences (details are given below). We restrict principals to selecting one of five investment strategies (very conservative, conservative income, balanced, growth, aggressive growth; following the Mutual Fund Dealers Association of Canada (2014)), rather than having them communicate an explicit investment share to their agents. Although strategies are ordered in an unambiguous way, there is variation over the exact interpretation of these verbal categories in terms of the investment share of the risky asset, allowing for "translation errors" in the communication between principal and agent. This allows principals to give agents the benefit of the doubt in case these do not

[30] Experiment 3 uses data of a larger study analyzing the behavior of financial advisors and clients in an advice relationship. While advisors' behavior is analyzed in a companion paper (Kling et al. 2019), the current paper focusses on the clients assessment of investment outcomes. The supplementary material reproduces the part of the instructions that is relevant for the decision process evaluation. Further details of the experiment and the instructions for other parts of the study are part of chapter two (Kling et al. 2019).

implement the strategy as perceived by the principal. We want to test how principals' satisfaction with the agents' investment decisions is influenced by the outcome of the investment, and in particular, whether good outcomes make it more acceptable that the agent did not follow the principal's request.

In the first part of the experiment, we ask participants for their perception of the five investment strategy terms. The task is to map the five individual strategies into investment shares (0% to 100% of total wealth) into a hypothetical risky asset. We keep the risky asset non-specific on purpose, as this is also a feature of real-life risk-classification terms. At the beginning of part two, participants learn about the Gneezy-Potters investment task. Participants consecutively play as both principals and agents (referred to as "client" and "financial advisor" in the experiment). Participants play as principals first and individually choose their preferred verbal investment strategy to be communicated to their agent. Subsequently, roles are switched and participants now take the financial investment decision as agents. On the decision screen, they are reminded of the structure of the risky asset, their compensation, and the principals' investment preferences. While adjusting the amount invested for the principal between 0 and 10€ (in steps of 0.10€, using a slider), they can observe a table of potential payoffs to their principal as well as to themselves which updates in real time. We vary two aspects in the decision by the agents. First, we vary whether an agent serves one or five principals. Second, we vary the incentive structure of the agent (fixed fee; participating only in the gains; participating to a limited degree in both the gains and the losses of the principal). We pool these conditions in the current analysis as they regard the agents' rather than the principals' behavior.

After all decisions have been made, the roles (principal or agent) are randomly selected to determine financial payoffs for each participant. Principals then see the following information: (i) the (verbal) investment level they demanded; (ii) the actual investment made by their agent; (iii) whether their investment was successful or unsuccessful; and (iv) their payoff. Importantly thus, they are prominently reminded of their preferred strategy, inducing a strong demand for judging on the basis of whether the agent implemented the request of the principal. Principals then pick one of four pre-determined messages to indicate their dissatisfaction or satisfaction with their agents' investment decision: "I am [very dissatisfied / dissatisfied / satisfied / very satisfied] with your investment decision." Finally, agents learn about the investment results and receive the message sent by their respective principals. The experiment concludes with a short questionnaire on demographics.

3.4.2 Setting and Summary Statistics

Experiment 3 was conducted at AWI-Lab, Heidelberg (GER). The experiment was programmed in oTree (Chen et al. 2016). Sessions and the participant pool were managed with hroot (Bock et al. 2014). The research question was not revealed to participants neither as part of the invitation to partake nor as part of the experiment itself. Each session lasted about 45 minutes and participants earned an average of €11.85.

In total, 324 participants took part in the experiment, yielding 162 observations. The average age of our participants at the time of the experiment was 23 years, 56% were female, and 29% indicated to be studying economics. Approximately 3% were studying psychology.

3.4.3 Results and Discussion

Table 3.5 shows satisfaction ratings of principals with their agents' investment decision. We say an agent *follows* the principal's wish if the invested amount falls into the range of investment shares that the principal associated with the communicated investment strategy.[31] We observe that principals are significantly more satisfied with the investment decision if the result is favorable than if it is unfavorable, replicating the relevance of outcomes. As we would expect, principals are also more satisfied if the agent implements their desired investment ("followed") than if she did not. However, this effect is only significant if the outcome was unfavorable. That is, for favorable investment outcomes, we do not observe a significant effect of the desired investment strategy anymore; the investment outcome moderates the effect of whether the agent implemented the principal's request. Quite strikingly, testing differences along the diagonal reveals that a decision which is in line with the principal's preference but results in an unfavorable random outcome is even seen as significantly less satisfactory than a decision which is at odds with the principal's explicit wish but by mere chance resulted in a favorable outcome (two-sample, two-sided Mann-Whitney-U test, average ratings 0.6 vs. 1.1, $z = -2.37$, $p < .05$). We also observe that even in the worst case of an unfavorable outcome

[31] While we can cleanly separate our principals into these categories for the analysis, we did not explicitly tell participants whether the investment was compatible with the stated preference in the experiment. Instead, principals had to infer whether the agent followed their wish from the observed investment amount.

when the agent did not follow, we only obtain a neutral assessment of -0.2 (z = -0.97, p = 0.33, Wilcoxon test). This observation is consistent with the positivity bias we also document in Experiment 2.

A multivariate analysis confirms the initial observations (Appendix 3B). We regressed investment satisfaction on indicators for agents following the principals' wishes and for observing favorable outcomes. Investment satisfaction, expressed through the messages sent to agents, is significantly positively affected by both agents following the principals' wishes as well as observing a favorable outcome of the random draw. Testing the linear hypothesis of equality of the respective coefficients, we can conclude that the effect on decision satisfaction of observing a favorable outcome is stronger than the effect of recognizing that an agent behaved in the principal's interest.[32] The results are unaffected by the inclusion of demographic control variables.

Table 3.5: Satisfaction with Decision

		Investment Profile		
		Followed	Not Followed	Difference
Outcome	Favorable	1.6	1.1	0.5
	Unfavorable	0.6	-0.2	0.8**
	Difference	1.0***	1.3***	

Notes: The table shows the average satisfaction of principals with the investment decision of their financial agent separated by the agent following or not following the principal's wish and by random outcome. The scale is -2 to +2 for very dissatisfied to very satisfied with the decision. Statistical significance is based on one-sided, two-sample Mann-Whitney-U tests. */**/*** denote 5%/1%/0.1% significance levels.
Favorable: Followed (22) vs. Not Followed (26), z = -1.68, p = .092.
Unfavorable: Followed (49) vs. Not Followed (65), z = -2.72, p < .01.
Followed: Favorable (22) vs. Unfavorable (49), z = -4.12, p < .001.
Not Followed: Favorable (26) vs. Unfavorable (65), z = -3.79, p < .001.

It is worth recalling how this experiment differs from the previous two studies. First of all, the invested amounts as well as outcome information are given to all principals. There cannot be any effects of having both outcome and process information versus process information only. Any outcome related effect must stem directly from the realizations of the random outcome draw. Principals evaluate agents who behave in line

[32] Wald test on the equality of both coefficients in the preferred ordered logit specification: model 3: chi2(1) = 5.08, p < .05; model 4: chi2(1) = 4.73, p < .05. The results are qualitatively similar but marginally insignificant in OLS regressions instead and testing the linear hypothesis with a two-sided F-test: model 1: F(1, 159) = 3.24, p = .074; model 2: F(1, 156) = 3.18, p = .077.

with their preferences significantly better than those who do not if the outcome was unfavorable. Given their ability to thus identify the relevant benchmark for evaluation, it is even more surprising that their decision process evaluations are even more strongly affected by the arbitrary (because random) outcome information.

Another aspect in which this experiment differs from our first laboratory experiment is the fact that the evaluation takes place in form of costless messages, rather than payoff-affecting rewards. Thus, social payoff considerations in the sense that gains from positive random outcomes could be shared between the two participants cannot play a role in the decision process evaluation. As it is a one-shot decision, the costless messages also cannot affect subsequent behavior or instill a "team spirit" between the two participants. This would only be possible in Experiment 1, with its dynamic multi-round setting and fixed principal-agent pairs. Ruling out social payoff considerations as well as concerns for the behavior in future rounds narrows down the number of alternative explanations for the effects at play.

3.5 Conclusion

We observe a clear outcome bias in principals' evaluations and rewards for financial agents in risky investment decisions. The outcome focus seems normatively questionable because it rewards lucky behavior on the basis of hindsight, rather than to reward good decisions on the basis of the information available to the agent. Importantly, it exists in settings where the decision process is clear and observable, and therefore there is no need to draw inferences about the decision from the outcome, as would be the case in situations with asymmetric information. In contrast to previous studies in the context of CEO salaries that have observed financial rewards for luck only if principals are weak (Bertrand and Mullainathan 2001), in the current experiment the effect was fully due to the principals' decision-making.

Studying the potential processes lying behind this outcome focus, we found that social preference effects, which may also loom large in situations outside the lab, might be a relevant aspect. Contingent on an outcome-based trigger to reward (random) successes, social comparison may play a role in defining the size of the rewards. However, outcome bias is relevant also in the absence of social comparison as shown in Experiment 2. Moreover, outcome bias seems more pronounced after good outcomes than after bad ones. This suggests that justification is an important aspect, and with either the decision or the outcome having a stronger influence depending on which turns out more justifiable. In contrast to Gurdal et al.'s (2013) interpretation, blame might not be the main driver of outcome bias in situations of (financial) agency. Our results also provide an interesting exception to the often-observed negativity bias (Baumeister et al. 2001).

Additional channels for the occurrence of outcome biases in the current experiments exist. The observed outcome-biased behavior may derive from the fact that in many situations outcomes are indicative of information available to the decision maker but not to the evaluator (Hershey and Baron 1992), or potentially provide the only available basis for judgments of the decision process (Baron and Hershey 1988). Consequently, a focus on outcomes may be inappropriately transferred to situations in which more or even all relevant information on the decision process is available. In the context of financial decisions with symmetric information about the investment and a large random component in outcomes, it is important for principals to understand and prevent outcome bias. Future research may fruitfully focus on the information formats that reduce outcome bias in financial agency.

Appendix 3A

This appendix derives the optimal behavior of an expected-utility maximizing agent with fairness preferences of the form proposed by Fehr and Schmidt (1999) for the case of outcomes (REWARD AFTER), and of the form proposed by Trautmann (2009) for the case of expected outcomes (REWARD BEFORE).

The general two-player variant of the utility function of player i in the presence of comparison to player j in the Fehr and Schmidt model is given by

$$U_i(x_i, x_j) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}, \qquad i \neq j,$$

where x_i and x_j denote the payoffs for each player and α_i and β_i denote the individual's parameters of inequity aversion. It is assumed that players suffer more from disadvantageous inequality than from advantageous ($\alpha_i \geq \beta_i$) and that players do not like to be better off than others ($0 \leq \beta_i < 1$).

In the case outcomes are observable (REWARD AFTER) and turn out favorable, the payoffs to the principal (x_P) and the agent (x_A) are given by

$$x_P = 100 - R + (100 - I) + 3.5I = 200 + 2.5I - R,$$
$$x_A = 100 + R,$$

where R and I denote the reward paid to the agent and the amount invested in the risky asset by the agent, respectively. The principal maximizes her utility

$$U_P(R) = 200 + 2.5I - R - \alpha_P \max\{-100 + 2R - 2.5I, 0\}$$
$$- \beta_P \max\{100 - 2R + 2.5I, 0\}$$

by choosing the reward $R \in [0,100]$ optimally. The resulting expected utility maximizing rewards are shown in the upper panel of Figure A1. They crucially depend on the level of investment in the risky asset and the parameter of advantageous inequity aversion.

In the case where outcomes are observable but turn out unfavorable, the payoffs for principals and agents equal:

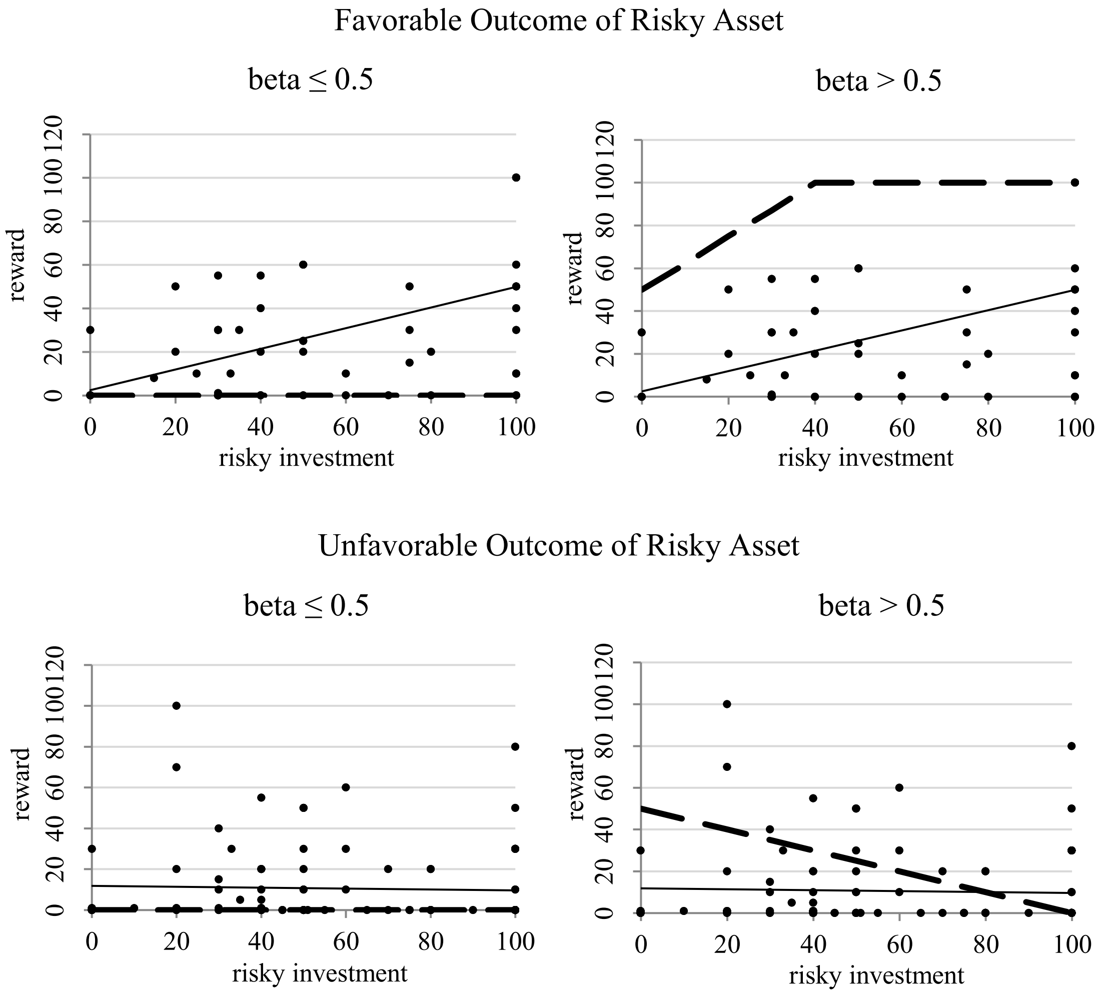
$$x_P = 100 - R + (100 - I) = 200 - R - I$$
$$x_A = 100 + R$$

As a result, the utility function of the principal becomes

$$U_P(R) = 200 - R - I - \alpha_P \max\{-100 + 2R + I, 0\} - \beta_P \max\{100 - 2R - I, 0\}.$$

The principal maximizes her utility by choosing the reward optimally. The expected utility maximizing rewards again depend on the parameter of advantageous inequity aversion and the risky investment by the agent. They are graphically illustrated in the lower panel of Figure 3A.1.

Figure 3A.1: Fehr and Schmidt's Outcome Fairness, REWARD AFTER



Notes: Scattered observations; linearly fit – solid line; EU-maximizing reward - dashed line.

Figure 3A.1 shows the qualitative predictions of the model in terms of expected utility maximizing rewards for modest (beta ≤ 0.5) and strong (beta > 0.5) inequality aversion. Clearly, the model cannot explain the observed reward pattern in its strict form assuming the same beta parameter for all participants. If we allow for a heterogeneous distribution of beta parameters, a different distribution of parameters is needed for the case of a successful investment versus an unsuccessful investment. For the favorable outcome, subjects should predominantly have large betas > 0.5. In contrast, for the unfavorable outcome prediction to match the data, subjects should hold small betas ≤ 0.5.

Trautmann's (2009) model of expected outcome fairness modifies the Fehr and Schmidt model by replacing the comparisons of realized outcomes with comparisons of expected outcomes. The general utility function for player *i* in the presence of comparison to player *j* in the two-player case is given by

$$U_i(x_i, x_j) = x_i - \alpha_i \max\{E[x_j] - E[x_i],0\} - \beta_i \max\{E[x_i] - E[x_j],0\}, \qquad i \neq j.$$

The assumptions about α_i and β_i remain unchanged.

In treatment REWARD BEFORE, only the amount invested in the risky asset is known to the principal at the time she chooses the reward for the agent. Consequently, she does not know her realized payoff and thus chooses the reward to maximize expected utility based on expected payoffs. The expected payoffs for the principal and the agent are given by:

$$E[x_P] = 100 - R + (100 - I) + \frac{1}{3} * 3.5 * I = 200 + \frac{1}{6}I - R$$

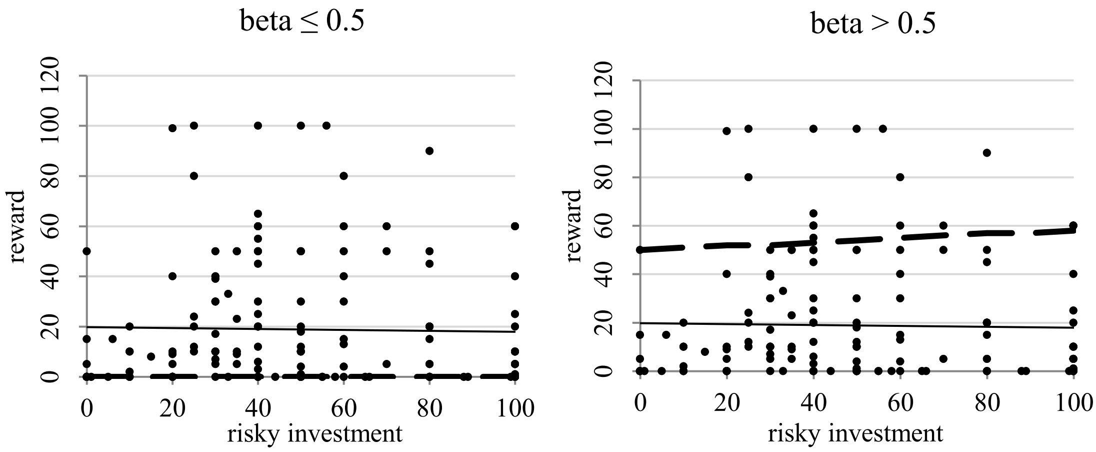
$$E[x_A] = 100 + R$$

Accordingly, the principal maximizes the utility function

$$U_P(R) = 200 + \frac{1}{6}I - R - \alpha_P \max\left\{-100 + 2R + \frac{1}{6}I, 0\right\} - \beta_P \max\left\{100 - 2R + \frac{1}{6}I, 0\right\}$$

by choosing the reward optimally. The resulting expected utility maximizing rewards depend on the risky investment by the agent and the principal's parameter of advantageous inequity aversion. Predictions are shown in Figure 3A.2. While the strict form of the model with a unique beta parameter for all principals cannot match the data, assuming a distribution of betas with roughly half of the participants below and above the 0.5 threshold would lead to predictions similar to the actual behavior

Figure 3A.2: Trautmann’s Expected Outcome Fairness, REWARD BEFORE



Notes: Scattered observations; linearly fit − solid line; EU-maximizing reward − dashed line.

Appendix 3B

Table 3B.1: Summary Statistics

Variable	Experiment 1	Experiment 2	Experiment 3
Age (SD)	22.5 (2.8)	38.9 (38.9)	23.3
Female	37.3%	47.5%	55.6%
Dutch	36.2%	-	-
Field of Studies			
Economics	55.2%	2.7%	29.0%
Business	36.6%	13.5%	0.9%
Psychology	1.5%	4.4%	2.8%
Law	3.0%	0.7%	1.8%
Other	3.7%	78.8%	65.4%
Participants	134	297	324
Agents	67	0	324
Principals	67	297	162

Notes: Experiment 1 was run in the Netherlands at a university with a large share of foreign students. We did not collect nationality information aside from asking whether participants were Dutch or not. Experiment 2 was run on Amazon Mechanical Turk and was restricted to participants located in the United States of America. Experiment 3 was run in Germany and we did not collect nationality information. In Experiment 1, half of the participants played in the role of agents, the other half in the role of principals. In Experiment 2 everyone took part in the role of the principal. In Experiment 3, participants took on both roles. Everyone made an investment decision as an agent. As payoff relevant roles were determined randomly before principals sent their messages expressing satisfaction with the investment decision, the number of observations is reduced to 162.

Table 3B.2: Experiment 2 - Multivariate Analysis of Outcome Satisfaction

	OLS		**Ordered Logit**	
Low investment	Outcome satisfaction		Outcome satisfaction	
Favorable outcome	2.8650 *** (0.2723)	2.9308 *** (0.2659)	3.7128 *** (0.5403)	4.1574 *** (0.5717)
Controls	No	Yes	No	Yes
# observations	98	98	98	98
High investment	Outcome satisfaction		Outcome satisfaction	
Favorable outcome	5.2363 *** (0.2419)	5.2752 *** (0.2430)	5.8998 *** (0.8514)	6.1119 *** (0.8907)
Controls	No	Yes	No	Yes
# observations	97	97	97	97

Notes: Base category is unfavorable outcome; standard errors in parentheses; *** denotes significance at the 0.1% level; controls are age, gender, education level and being an economist.

Table 3B.3: Experiment 2 - Multivariate Analysis of Decision Satisfaction

	OLS		Ordered Logit	
Low investment	Process satisfaction		Process satisfaction	
Unfavorable outcome	0.1495 (0.3211)	0.0831 (0.3179)	0.0759 (0.3559)	0.0106 (0.3607)
Favorable outcome	1.1878 *** (0.3178)	1.0232 ** (0.3158)	1.3982 *** (0.3755)	1.2846 ** (0.3792)
Controls	No	Yes	No	Yes
# observations	149	149	149	149
High investment	Process satisfaction		Process satisfaction	
Unfavorable outcome	-0.6575 * (0.3288)	-0.7194 * (0.3352)	-1.1768 ** (0.3951)	-1.3091 ** (0.4049)
Favorable outcome	2.4620 *** (0.3305)	2.4190 *** (0.3358)	2.2869 *** (0.4006)	2.2573 *** (0.4031)
Controls	No	Yes	No	Yes
# observations	148	148	148	148

Notes: Base category is unknown outcome; standard errors in parentheses; */**/*** denote significance at the 5%/1%/0.1% level; controls are age, gender, education level and being an economist.

Table 3B.4: Experiment 3 - Multivariate Analysis of Decision Satisfaction

	OLS		Ordered Logit	
	1	2	3	4
Follow	0.6614 ** (0.1870)	0.6526 ** (0.1876)	0.9724 ** (0.3060)	0.9779 ** (0.3095)
Good Outcome	1.1654 *** (0.2032)	1.1524 *** (0.2041)	1.9899 *** (0.3642)	1.9687 *** (0.3661)
Controls	No	Yes	No	Yes
# observations	162	162	162	162

Notes: Ordered logistic regressions; coefficients reported; standard errors in parentheses; The dependent variable is satisfaction with the investment decision from messages sent by principals to agents. **/*** denote significance at the 1%/0.1% level.

Chapter 4

Ambiguity Attitudes in Decisions for Others

Abstract. We probe the pattern of ambiguity aversion for moderate-likelihood gain prospects, and ambiguity seeking for low-likelihood gain prospects, if people make decisions not for themselves but as agents for others. We confirm the pattern both with and without accountability.[33]

[33] This chapter is co-authored by Stefan Trautmann and published as König-Kersting and Trautmann (2016).

4.1 Introduction

Experimental research on decision making under risk has found marked differences between decisions for oneself and for others (e.g., Chakravarty et al. 2011, Füllbrunn and Luhan 2017), with accountability being suggested as a moderating factor (Pollmann et al. 2014). Building on this research, we observe that decisions under uncertainty are often characterized by a lack of knowledge about the probabilities attached to the various outcomes. In contrast to decisions under *risk* (where probabilities are known), these decisions are referred to as decisions under *ambiguity*. We study whether the pronounced self-other disparities observed under risk also emerge for decisions involving ambiguity. Given the close similarity to risky decision settings, agency and accountability would be expected to have effects for ambiguity as well.

Previous literature has found a complex pattern of attitudes toward ambiguity, with people being ambiguity averse for moderate likelihood gains (as in the classic Ellsberg 2-color task) and ambiguity seeking for low likelihood gains.[34] Studying decisions for others with and without accountability, we probe the robustness of this pattern outside the context of individual decision-making.

The next section describes the experimental setup. The following section presents the results, showing that the pattern of ambiguity aversion and seeking suggested in the previous literature emerges strongly in both decisions for oneself and for others. We do not observe self-other disparities for ambiguity attitudes. The final section discusses these findings in the context of the related literature.

4.2 Experimental Design

4.2.1 Decision Tasks

We measure ambiguity attitudes using Ellsberg-urn tasks with either 2-color urns (moderate likelihood) or 10-color urns (low likelihood) as described in Trautmann and van de Kuilen (2015). We implement these two settings in a between-subjects design. In both settings, participants choose between betting on a red chip drawn from 100-chip bag with a known distribution of colors (risky prospect), and betting on a color of their choice

[34] See e.g. Trautmann and van de Kuilen (2015) for a recent review of the literature. The reverse pattern has been observed in the loss domain. Given the complexity of the agency setting, we focus on the gain domain in the current paper, avoiding issues of implementing losses.

from a 100-chip bag with an unknown distribution of colors (ambiguous prospect).[35] A successful bet yields a prize of €10; otherwise, the payoff equals €0. In each setting, participants make seven choices between risky prospects with varying number of red chips, and the ambiguous prospect. In the moderate likelihood task, the bags contain red and blue chips. In the low likelihood task, the bags contain chips of 10 different colors. The seven choices are presented sequentially on separate screens, always starting with the ambiguity-neutral risky prospect. In the 2-color task, this bag contains exactly 50 red and 50 blue chips. In the 10-color task, it contains 10 red and 90 chips of different color. The seven choices for each task are shown Table 4.1, in the order they were presented to the participants. Note that our elicitation method makes preference consistency requirements much less salient than commonly used single-screen choice lists with items presented in ascending order.

Table 4.1: Order of Decision Problems and Winning Probabilities
for the Risky Prospect

Decision	2-color task: Winning probability of risky prospect vs. 2-color ambiguous bet	10-color task: Winning probability of risky prospect vs. 10-color ambiguous bet
1	0.50	0.10
2	0.35	0.01
3	0.65	0.19
4	0.40	0.04
5	0.60	0.16
6	0.45	0.07
7	0.55	0.13

Our setup allows us to collect two pieces of information. First, decision 1 allows us to determine ambiguity attitudes as typically done in single-choice tasks, unaffected by any considerations of order or choice-list effects. Together with decisions 2 to 7, we are then able to determine a *probability equivalent* (PE) for the ambiguous prospect, defined as the mid-point between the lowest risky probability for which the decision maker

[35] Having participants choose their own winning color prevents the ambiguous bags from being strategically filled to the participants' disadvantage. This problem does not obtain for the known-distribution risky bags. For practical reasons thus, participants cannot choose their winning colors for the risky prospects, because it would require a large number of additional bags to cover all possible color choices and chip distributions.

chooses risky and the highest risky probability for she chooses ambiguous. For example, if the decision maker chooses risky in the 2-color task except for risky probabilities 0.35 and 0.40, we calculate her PE as 0.425.[36,37] In both tasks, a lower PE indicates lower tolerance of ambiguity. In contrast to the simple initial choice, the PE provides a more fine-grained measure of attitude with more variation and statistical power.

4.2.2 Treatment Manipulation

We implement the above-described decision task in three treatments. In treatment SELF, participants make the decision for their own account. In treatment OTHER, they (agent) make the decision for another participant (principal) who remains passive in the task. In treatment REWARD, they make the decision for another participant who is then asked if she wants to reward the decision maker for her choice. All participants in all treatments receive a fixed payment of €2 (on top of a show-up fee of €3). In treatment REWARD, the principal can use this amount to transfer a reward to the agent. Specifically, one of the seven decisions made by the agent is randomly selected and implemented for real. The principal observes the choice and the outcome, and is then asked which amount between €0 and €2 she wants to transfer to the agent (in increments of €0.20). Both principals and agents know the procedure and the available amounts. Thus, agents can anticipate the effect of their choice on their potential rewards.

In conditions OTHER and REWARD, half the participants make choices as agents in the 2-color task and the other half make choices as agents in the 10-color task. Subsequently, each agent serves a principal in the other task. That is, the initial choice behavior is not affected by their later experience as a principal. Participants learn about the details of the other task only after they made choices in the task in which they act as agents. One of the two settings was selected for payment.[38]

[36] For some participants the PE is not defined by the procedure because they choose ambiguous for risky probability p and risky for risky probability q, with q<p. In this case we define an indifference range, i.e. we consider the risky probability at which the participant first switches from the ambiguous to the risky prospect and the last choice item at which the participant switches back from the risky to the ambiguous prospect. We then define the participant's probability equivalent as the midpoint of this indifference range. Our results remain unchanged if we only use observations with a single switching point instead.

[37] For participants who always choose risky (ambiguous), we define the PE as 0.325 and 0 (0.675 and 0.205) in the 2-color and the 10-color tasks, respectively.

[38] In condition SELF, for participants in the 2-color (10-color) task, the second part of the experiment had them make choices in the 10-color (2-color) task, to keep the two-part structure equivalent to the agency condition. We only use the initial (between-subjects) choices that are unaffected by order effects to keep the structure identical to the agency condition. Our results remain qualitatively unchanged if we use all choices of treatment SELF instead.

4.2.3 Lab Procedures

The experiment was programmed using z-Tree (Fischbacher 2007). Participants were invited from the participant pool using hroot (Bock et al. 2014). All participants received a show-up fee of €3, a fixed payment of €2, and could earn €10 from the choice task. The experiment took about 40 minutes, for which participants earned €7.16 on average.

Before each session, the ambiguous bags were filled anew with 100 chips, which were drawn from larger bags containing 100 chips of each color. The risky bags were checked to contain the correct distributions of colored chips. The physical bags were visibly placed on the experimenter's table and could be inspected by participants after each session.

After all choices were made, one setting (2-color or 10-color task) was randomly selected, and uncertainty was then resolved by drawing chips physically with the help of a volunteering participant. Results were entered into the program. For each participant or each agent-principal group, the computer randomly selected one choice problem and calculated payoffs[39]. In REWARD, at this point the principals learned about the decision by the agent and their outcome, and made their decision about the reward. Final payoffs were calculated; participants answered a demographic questionnaire, were paid and dismissed from the lab.[40]

4.3 Results

In total, 194 student subjects participated in the experiment (SELF: 38, OTHER: 78, REWARD 78), of which 47.9% (93) are female, and 36.6% (71) are economics students. Consistency in the two choice tasks is high, given that decisions are not presented in ascending order of probability and are shown on separate screens. In the 2-color task, 73.3% (85) of the decision makers were consistent and for 85.3% (99) we could calculate PEs. In the 10-color task, consistency is even higher at 81.9% (95), with PEs being calculated for 92.2% (107).

[39] Selecting one of the Ellsberg-type deicisions for payment can be problematic as it opens the opportunity for unintended hedging (Oechssler and Roomets 2014). Yet, making all decisions payoff relevant has its own drawbacks as well. These include income and wealth effects. Azrieli et al. (2018) conclude that paying one randomly selected problem may be the best choice afterall.

[40] All files necessary for replicating the experiment and the results are available at https://heidata.uni-heidelberg.de/dvn/dv/awiexeco .

We first look at choice behavior in the initial choice item in each task. This decision problem requires subjects to choose between the ambiguous prospect and a risky prospect, which is equivalent to the ambiguous prospect under expected utility. If participants are ambiguity neutral, we expect on average one half or the decision makers to choose the risky and the other half the ambiguous prospect randomly. The upper panel of Table 4.2 shows the percentages of ambiguous choices in these decision items for each of the three treatments in both tasks. For the moderate-likelihood task, we find a tendency towards ambiguity aversion that is marginally significant in SELF and OTHER, and insignificant in REWARD (p = 0.1996, binomial-test). For the low likelihood task, we find insignificant ambiguity seeking for SELF (p = 0.1671, binomial-test), marginally significant ambiguity seeking for OTHER and significant ambiguity seeking for REWARD.[41] Note that our results replicate the gain domain parts of the fourfold pattern of ambiguity attitudes reported by Trautmann and van de Kuilen (2015).

Table 4.2: Ambiguous Choices in First Decision Item and Probability Equivalents

	SELF		OTHER		REWARD	
Binary choice percentages						
Moderate likelihood (2-color task)	26.3%	AA*	33.3%	AA*	38.4%	AA^{ns}
Low likelihood (10-color task)	68.4%	AS^{ns}	66.7%	AS*	74.4%	AS***
Median probability equivalents (means in parentheses)						
Moderate likelihood (2-color task)	0.475 (0.442)	AA**	0.475 (0.468)	AA**	0.475 (0.466)	AA***
Low likelihood (10-color task)	0.115 (0.112)	AS**	0.115 (0.136)	AS***	0.115 (0.123)	AS***

Notes: Choice entries give % of ambiguous prospects chosen; Binary choice: two-sided binomial test against $p = 0.5$; Probability equivalents: Wilcoxon-test against 0.5 / 0.1; *, **, *** denote significance at 10%, 5%, 1%. Direction of effect: AA = ambiguity averse; AS = ambiguity seeking.

Our second research question concerns the comparison of ambiguity attitudes across treatments. We find no significant treatment differences in the initial choice items for

[41] Dealing with the first decision items only, this analysis includes all observations. If we restrict it to observations fulfilling consistency over the set of choices, the pattern becomes more significant.

either task.[42] The lower panel of Table 4.2 shows probability equivalents for the three treatments and both tasks. The pattern found in the initial binary choice is confirmed by the probability equivalents, but results are more significant: In the moderate likelihood task, we find significant ambiguity aversion, and in the low likelihood task, we find significant ambiguity seeking, in all conditions. Comparing across treatments, we do not observe significant differences in either task.[43]

4.4 Discussion and Conclusion

We set out to study whether ambiguity attitudes are affected by agency situations and if so, how they change. Recalling the substantial differences reported for decision making under risk in very similar settings, our main finding that participants' ambiguity attitudes are unaffected by the agency setting is quite unexpected. In our experiment, participants show equally pronounced attitudes in the agency conditions as in decisions for their own account.

Chakravarty et al. (2011) and others have found that people make more risk-neutral decisions as agents. Similarly, Charness at al. (2013) suggested that ambiguity neutrality might be more normatively compelling in social settings (group decisions in their study). In contrast, we do not observe a tendency of agents to make more ambiguity neutral decisions: strong ambiguity aversion for moderate likelihood prospects and ambiguity seeking for low likelihood prospects emerge under agency. We infer that ambiguity attitudes are more robust than risk attitudes with regard to social interactions and peer effects. Indeed, Trautmann et al. (2008) find that being observed by others leads to stronger ambiguity aversion, questioning the normative appeal of ambiguity neutrality in social settings.

In line with recent studies, we find that ambiguity attitudes depend strongly on the likelihood range considered. Despite the fact that our elicitation method differed from previous experiments and that in some conditions the decisions are made for others, we

[42] Initial choice items, moderate likelihood task: SELF vs. OTHER $p = 0.764$; SELF vs. REWARD $p = 0.397$; OTHER vs. REWARD $p = 0.814$. Low likelihood task: SELF vs. OTHER $p = 1$; SELF vs. REWARD $p = 0.756$; OTHER vs. REWARD $p = 0.620$; Fischer's exact tests.

[43] Probability equivalents, moderate likelihood task: SELF vs. OTHER $p=0.398$; SELF vs. REWARD $p=0.488$; OTHER vs. REWARD $p=0.920$. Low likelihood task: SELF vs. OTHER $p=0.189$; SELF vs. REWARD $p=0.371$; OTHER vs. REWARD $p=0.365$; Mann-Whitney tests.

replicate the fourfold pattern of ambiguity attitudes in the gain domain. The result supports the external validity of the pattern of ambiguity attitudes.

Chapter 5

Bank Instability:
Interbank Linkages and the Role of Disclosure

Abstract. We study the impact of disclosure about bank fundamentals and interbank linkages on depositors' behavior. Using a controlled laboratory environment, we identify under which conditions disclosure is conducive to bank stability. We find that bank deposits are sensitive to perceived bank performance. While banks with strong fundamentals benefit from more precise disclosure, an opposing effect is present for banks with poor fundamentals. Our findings highlight both the costs and benefits of bank transparency and suggest that disclosure is not always stability enhancing.[44]

[44] This chapter is co-authored by Stefan T. Trautmann and Razvan Vlahu.

5.1 Introduction

In the aftermath of the financial crisis, greater regulation and efforts to increase the transparency of the banking industry have been at the forefront of the policy debate. Rigorous stress testing has been introduced as a key method of assessing the financial sector's ability to withstand large-scale correlated shocks to multiple (macro-)economic factors. With the rise of these regular tests of risk-bearing ability and capital adequacy of financial institutions on both sides of the Atlantic, the questions of whether or not to release results publicly and at what level of detail, have been discussed controversially by politicians, researchers and media outlets alike.

The reason for the observed controversy can be understood by looking at the trade-off between market discipline and financial stability. On the one hand, it is clearly in depositors' and investors' interest to know the state of their financial institutions in order to be able to make well-informed financial decisions. Increased public awareness of bank risks may thus enhance market discipline, which penalizes financial institutions for excessive risk taking. At the same time, it is also clear that insolvent financial institutions need to be identified and resolved quickly in order to prevent subsequent negative ripple effects on other institutions, potentially endangering the whole banking system. As evidence from the recent crisis suggest, uncertainty about which banks incurred losses may lead to situations in which banks are unable to raise additional funds to withstand liquidity demand because of market freeze (i.e., potential lenders were unable to assess banks' solvency due to balance sheet opacity, and as a result, fearing information asymmetries, they were reluctant to lend). On the other hand, disclosing stress test results to the public may also have self-fulfilling effects in the sense that knowledge of an institution's subpar, yet not in itself dangerous result, may still lead to strong depositor reactions and a dramatic tightening of liquidity. Such liquidity squeeze might then lead to a bank failure, regardless if the bank is solvent in the long run or not.

Furthermore, stress tests usually cover only a subsample of all financial institutions, leaving depositors of untested banks in the dark even if results are published for others. This aspect highlights the potential importance of knowledge about economic linkages between financial institutions. How similar are different banks in their capital adequacy? Are various banks exposed to the same levels and types of risk? Knowledge of these kinds of economic linkages in the financial sector can be crucial in understanding if and how disclosed information about certain institutions may lead to panic behavior among depositors with the potential to subsequently spread to other institutions in a contagious fashion.

In the current study, we focus on the direct information-based mechanism. Taking this perspective, we argue that factual information of varying precision is gathered by depositors and passed on to others by means of communication, rather than observation of actual behavior. Real word justification for this approach can be found in the stylized sequence of events in bank runs. Large reductions of bank deposits through wire transfers often preceded the more easily observable depositor run at bank counters. Statistical information about deposit levels are usually published with a lag of multiple months, precluding timely observation of withdrawals through channels other than actual cash withdrawals at ATMs and counters. One of the most recent examples of a depositor run following this sequence is Greece, where deposit levels have fallen tremendously after the elections of 2014, yet the more easily observable depositor run by retail customers only started about half a year later (European Central Bank 2015).

The theoretical literature provides useful insights on the underlying mechanisms of bank runs, information disclosure and contagion effects. However, there is not much empirical work on the potential effects that information precision about bank fundamentals, as well as the simultaneous consideration of both disclosure about fundamentals of individual banks and information about economic linkages across banks, might have on depositors' behavior and thus on financial stability. We study these fundamental mechanisms in a laboratory experiment. This approach allows us to implement tighter control over the decision situation and cleaner treatment manipulation than would be possible by basing the analysis on empirical data and natural experiments. At the same time, it offers us the opportunity to study the effects of information disclosure on depositors' behavior in the absence and presence of economic linkages between financial institutions in a unified setting. Our setting is based on the Diamond and Dybvig (1983) framework, which treats bank runs as *coordination games* with inherent strategic uncertainty. In this set-up, we examine first, how different degrees of information precision about a bank's fundamentals create conditions for bank runs, and second, how noisy information about interbank linkages in combination with transparency over the fundamentals of one bank may trigger a run at another bank for which there is no disclosure.

Our paper fits into several strands of literature. First, it is related to the literature examining the effects of information disclosure and, more specifically, the debate on the publication of bank stress test results. Second, it is linked to the bank run literature in general, and to the experimental bank run literature in particular. We review the relevant literature in section 5.2. Section 5.3 presents the stylized banking setting for our experiment and introduces the depositors' coordination problem. Section 5.4 describes the experimental design and procedures. Section 5.5 formulates our hypotheses. Results are presented in Section 5.6 whereas section 5.7 concludes the paper.

5.2 Related Literature

5.2.1 Financial Disclosure

Morris and Shin (2002) highlight the potential for adverse effects of publicly releasing information. They argue that information might be "too effective" (p. 1522) in influencing behavior of market participants, as they tend to overreact to the information provided. Thus, even noise might affect behavior and worsen outcomes for market participants. They show that agents do not necessarily have to act irrationally for these effects to arise.

In similar vein, Nier (2005) starts from the idea that disclosure can be a bad thing as it might aggravate the situation at hand. However, he concludes that the net effect of transparency is a reduction in severe banking problems and an enhancement of financial stability. Nier and Baumann (2006) add to these results by demonstrating that absent of governmental safety nets for financial institutions, information disclosure can strengthen market discipline and lead to larger capital buffers of banks. They find government support to be detrimental to the effectiveness of disclosure in enhancing financial stability.

Adding to the observation that the effects of information disclosure in financial settings is highly context dependent, Bouvard et al. (2015) find that disclosing bank-specific information enhances the stability of the financial system during crises, but has a contrary effect in normal times. Based on their theoretical model, they deduce that regulators should increase transparency during crises. However, as the authors are quick to highlight, this behavior signals a deterioration in economic fundamentals, which, if anticipated, constitutes an incentive not to disclose the information in the first place. Connectedly, Goldstein and Leitner (2018) attempt to formulate an optimal disclosure policy. Assuming a regulator that has information about banks' ability to overcome future liquidity shocks, they find that during times of distress, partial disclosure appears optimal, while in non-crisis times, not disclosing information is favorable.[45]

Apart from the more general literature on the disclosure of financial information, there are also articles directly concerned with the publication of stress test results. Goldstein and Sapra (2014) discuss the issue in the light of costs and benefits. They conclude that disclosure of stress test results promotes financial stability, even though it might come at a cost. Leitner (2014) generally adds to the argument, but highlights that banks with weaker fundamentals might suffer from increased disclosure and due to market participants overreacting to it (cf. Morris and Shin 2002).

5.2.2 Bank Run Experiments

Arifovic et al. (2013) model bank runs as phenomena of pure coordination failure. They systematically vary the coordination parameter and characterize three regions, which correspond to no-run, indeterminacy, and run situations. While behavior of depositors shows some path dependence, coordination outcomes are generally difficult to predict if the coordination parameter falls into the indeterminacy region. Building on this result, Arifovic and Jiang (2014) demonstrate the effectiveness of random public announcements as sunspot coordination devices. Depositors react most strongly to announcements in times of high uncertainty, which highlights the need for careful consideration of financial disclosure in times of economic crises.

Schotter and Yorulmazer (2009) focus on the dynamics and severity of bank runs, rather than their occurrence. They demonstrate experimentally that insiders who know more about the crises as it develops are less likely to withdraw than uninformed depositors are. Their results support the theoretical findings on the importance of information availability for depositors' behavior in bank run contexts. Further evidence on the contextual sensitivity of financial disclosure is provided by Davis and Reilly (2016) who find that its effects also depend "on the complexity the additional information adds to the strategic situation" (p. 1015). Adding to the evidence, Shakina and Angerer (2018) study depositors' behavior in a much less restricted setting than previous studies. Their depositors can continuously withdraw and re-deposit funds without any order being

[45] For a review of different channels through which financial disclosure works in financial markets, refer to Goldstein and Yang (2017).

enforced. They find information regarding the state of the economy to influence withdrawals, but also highlight the importance of coordination.

While factual information clearly is an important determinant of behavior, most experimental setups in the bank run context also feature an element of strategic uncertainty, i.e. uncertainty about the behavior of others which co-determines own outcomes. The presence of strategic uncertainty opens the door for other considerations. Garratt and Keister (2009) provide evidence for the role of beliefs concerning the behavior of other depositors in affecting own withdrawal behavior. Utilizing a global-game setting (cf. Carlsson and van Damme 1993), Heggelin (2015) focuses on the effects of previous experience, risk aversion, level-k thinking, and disclosure quality on sensitivity of investors to bad signals about bank fundamentals. While reiterating the evidence for larger noise in disclosed information having an increasing effect on the prevalence of banking crises, he also points to individual characteristics such as risk preferences and previous experience to shape behavior.[46]

5.2.3 Financial Contagion

Finally, our study is also linked to the issue of financial contagion. Iyer and Peydró (2011) study contagion between banks using evidence from a natural experiment in India. They report robust evidence on higher interbank exposure leading to larger deposit withdrawals. The strength of the contagion effect experienced by exposed banks depends on their fundamentals. Weaker banks face larger contagion effects. Their study nicely shows the joint relevance of information about banks' fundamentals and knowledge of the economic linkages for depositors' decisions.

Taking the research question to the experimental laboratory, Chakravarty et al. (2014) study causes of bank run contagion based on the Diamond and Dybvig framework. They find evidence for contagion between two banks, independent of their fundamentals being economically linked or not. Brown et al. (2017) also study experimental coordination games to gain an understanding of the information conditions that lead a panic-based depositor run at one bank to trigger a panic-based depositor run at another

[46] Similar effects on depositor characteristics are reported by Trautmann and Vlahu (2013) for negative past experiences and loss aversion. Klos and Sträter (2013) suggests the ability for sophisticated reasoning to play a role by demonstrating that depositors' behavior fits the predictions of level-k models. In this regard, Kiss et al. (2016) show cognitive abilities as measured by the cognitive reflection task to predict withdrawals in the presence of strategic uncertainty. Finally, Dijk (2017) specifically studies the effects of emotions on depositors' behavior. He finds that background fear significantly increases the likelihood of withdrawals.

one. They identify changes in beliefs triggered by observing a depositor run as the reason for making own withdrawals more likely. In contrast to Chakravarty et al. (2014), they only find evidence of contagion in the presence of economic linkages between financial institutions. The results of Brown et al. (2017) are supported by Cipriani et al. (2018) who consider the informational channel of financial contagion. They find evidence of contagion between two markets, but only as long as asset fundamentals are correlated. Participants only apply information across markets, if it is rational to do so.

Trevino (2019) uses a global games approach, which allows for both channels of contagion (informational and fundamental) to operate simultaneously. She finds that information is not extracted optimally and, as a result, participants underweight their prior in fashion of a base rate neglect, which weakens the fundamental channel compared to the theoretical predictions. Similarly, an overreaction bias is affects the social learning channel negatively. She finds that too much weight is put on the information of others, even if it is irrelevant.

Finally, Kaufman (1994) and Glasserman and Young (2016) review large parts of the relevant literature. The latter also highlight information contagion as a mechanism that can be triggered by changes in perceptions about the creditworthiness of institutions and the value of their assets. These changes can propagate through the financial system and result in a general crisis of confidence.

5.3 Banking Setting

Given the large variety in approaches to studying information disclosure and economic linkages in experimental bank run settings, it seems prudent to start with a general explanation of the particular banking setting we have in mind. In this paper, we study an economy with three dates (0, 1, 2) and no discounting. A bank operating in this economy takes deposits at date 0 and invests in assets that produce profits at date 2. Bank's deposits are uninsured and costly.[47] The creditors are repaid (with interest) at date 2 if their bank is solvent. Solvency depends on the bank's assets portfolio and depositors' actions. With

[47] The evidence on the link between deposit insurance and depositors' behavior is tenuous. Flannery (1998) finds that insured depositors are concerned about the solvency of their bank, as well as about that of deposits insurer. Deposit insurance schemes may not be credible (Martinez Peria and Schmukler 2001, Prean and Stix 2011), the coverage of the deposit insurance funds is limited (Demirguc-Kunt at al. 2005, 2015) and even insured depositors may withdraw deposits from distressed banks (Iyer and Puri 2012, Karas et al. 2013). Calomiris and Jaremski (2016) review the theoretical arguments behind the creation of deposit insurance and the empirical evidence on its performance.

respect to the former, we assume banks' fundamentals (e.g., liquidity position, quality of assets). With respect to the latter, depositors are facing uncertainty about the quality of banks' assets and may choose to withdraw their money before maturity, at date 1. In order to meet its payment obligations at date 1 the bank may be forced to liquidate (some of) its assets. Conditional on the liquidity and quality of bank‘s assets at date 1, liquidation may be possible at a substantial discount. When the discount is too large, the bank may run the risk of not being able to pay back the remaining depositors at date 2, effectively rendering the bank insolvent. In this case, the bank is liquidated at date 1 and the liquidation value of its assets is distributed among those depositors who chose to withdraw. Upon bank bankruptcy, patient depositors (i.e., those without withdrawal claims at date 1) lose their deposits.

Information about the banking system is conveyed to market participants through disclosure. There are two types of disclosure, which may affect bank stability in this framework. First, there is the transparency about the quality and liquidity of bank's assets, which is arguably of highest importance to market participants. Such enhanced information about the bank's exposure to potential liquidity shocks may prevent (or, conditional on the type of information conveyed to the market, precipitate) individual bank runs as well as contagion effects across banks. Naturally, this type of disclosure may vary in its informativeness to depositors. Specifically, as we discuss in detail in Section 5.4, we consider various scenarios in which disclosed information about the banks' ability to withstand liquidity shocks is either non-informative, partially informative or fully informative. We assume that disclosed information is common knowledge among all depositors of a bank. More explicitly, all depositors receive the same information at the same time and no depositor has an advantage over the other depositors in reacting to it.

Second, the quality of information about the interbank linkages may contribute to the fragility of the banking sector. Common assets exposure is one important form of interbank linkages (Chen 1999; Ahnert and Georg 2018). Our experimental design captures this specific form. There are other forms of interbank linkages (e.g., interbank lending) but we abstract from them in this paper. Depositors typically face uncertainty about the existence of such linkages across different financial institutions. At one extreme, depositors might face maximum uncertainty when they are not aware of any explicit interbank linkages between their bank and other banks in the system. Rationally, information disclosed about the capacity of another bank to withstand liquidity shocks is not informative about the liquidity position of their own bank. At the other extreme, depositors may be aware that their bank has an identical asset portfolio as other banks. In this case, information about one bank is informative about the fundamentals of other bank. In reality, the precision of information about the interbank linkages generates various potential scenarios between these two extreme cases. We deliberately abstract away from different aspects of similarity and instead model similarity as the probability of being identical. As we discuss in the next section, we consider various scenarios in which disclosed information about the interbank linkages is either non-informative or partially informative.

5.4 Experimental Design

5.4.1 Banks and Depositors

We model banks as one-shot, three-player coordination games with Pareto-ranked run and no-run equilibria in pure strategies. Each bank has three depositors who can individually choose between withdrawing and not withdrawing their money. All depositors act simultaneously and without knowing other depositors' decisions. To model banks with different risk exposures and to allow financial disclosure to provide meaningful information to depositors about bank fundamentals, we consider three types of banks: Good, Medium, and Weak. The banks differ with respect to their payoffs for depositors in case of early liquidation as well as in the case of no liquidation. Put differently, we consider stochastic returns on deposits (instead of a fixed payment) in order to capture the role of uncertainty about expected returns on deposits on withdrawal decision. Note that all the banks in our experiment are solvent. There is no exogenous shock to their asset portfolios and all the banks, regardless their type, are able to repay depositors in full if none of them withdraws before maturity.

Good banks have the strongest fundamentals. They are the least fragile to liquidity shocks and fail only if two or more of their depositors withdraw. These banks offer the highest payoffs to depositors regardless the number of withdrawals. If all depositors keep their money in the bank, the bank does not have to liquidate any investments and all depositors receive a payoff R_G. If one depositor withdraws, the bank is able to repay him R_{Gw}, with $R_{Gw}<R_G$, thus the early depositor forgoes some of the potential future return. When at least two depositors withdraw, the bank is liquidated and the liquidation value

L_G is shared among early depositors. In case of bank liquidation, the depositor (if any) who decides to keep money in the bank receives zero.

Medium banks are more fragile than good banks and fail if at least one depositor withdraws. In terms of payoffs, they are identical to good banks in case of no liquidation, i.e. when nobody withdraws each depositor receives R_G. However, they have a lower liquidation value L_M, with $L_M < L_G$. As with *Good* banks, in case of liquidation the depositors withdrawing from a failed bank share the available funds among themselves leaving nothing to the other depositors.

Finally, the *Weak* banks are identical with *Medium* banks in terms of fragility (i.e., they fail if at least one depositor withdraws) and payoffs upon liquidation (i.e., liquidation value is L_M). However, they are less profitable than *Medium* banks and therefore pay less to their depositors in case of no liquidation: R_W, with $R_W<R_G$.

Table 5.1 presents the payoff matrix for this three-person coordination game. The payoff structure can be rationalized as follows: Some banks may get exposure at date 0 to the same asset class (e.g., real estate). The individual bank's specific investments are not observable, though. Ex-ante, the banks have identical expected returns and face identical cost of funding. This is due to the fact that the market does not have detailed information about individual banks' portfolios, but only aggregate information about the sectors to which the banks' are investing in. However, after the investment is made and before the returns are realized, banks' depositors may receive some information about the quality of banks' assets. Upon receiving such information (via mandatory or voluntarily bank disclosure), depositors may find out that some banks have more valuable/liquid assets than other banks. For example, one bank may turn out to have a larger exposure to the prime real estate sector than another bank, which is heavily exposed to the subprime sector. This revelation may affect not only banks' valuation but also their perceived capacity to withstand depositors' withdrawals. Exposure to the subprime market may be associated with illiquidity: Banks investing in this real estate segment, when forced to liquidate their investments, are able to do so only at large discounts. This increases their vulnerability in face of depositors' demand for liquidity. Our payoff structure is motivated by the idea of capturing the role of disclosure in offering additional information to banks' depositors about the quality (and liquidity) of banks' assets at a certain point in time after the initial investment.

Table 5.1: Depositors' Payoff Structure

Bank type and own decision	Number of other depositors withdrawing 0	1	2
Good			
not withdraw	R_G	R_G	0
withdraw	R_{Gw}	$L_G/2$	$L_G/3$
Medium			
not withdraw	R_G	0	0
withdraw	L_M	$L_M/2$	$L_M/3$
Weak			
not withdraw	R_M	0	0
withdraw	L_M	$L_M/2$	$L_M/3$

Notes: Appendix 5A shows the actual payoff structure implemented in the experiment.

Importantly, with the payoffs used in our experiments (see also Table 5A.1 in Appendix 5A), all of the three resulting coordination games (for Good, Medium and Weak banks) have two Nash equilibria in pure strategies. In the better of the two equilibria, no depositor decides to withdraw their money from the bank and everyone enjoys the high payoffs at asset maturity. In the worse equilibrium, all depositors withdraw their money early and share the liquidation value.

5.4.2 Treatments

The aim of our study is to examine whether different degrees of information, and the simultaneous consideration of both disclosure about fundamentals of individual banks and information about interbank linkages, may affect depositors behavior and thus financial stability. The degree of disclosure about individual banks and interbank linkages varies between groups of participants. This variation allows us to observe the outcomes of their coordination games and to identify the conditions that make coordination failure (i.e., a bank run) most likely.

Individual bank disclosure

The first dimension of interest is disclosure about an individual bank (Bank A, hereafter). Participants take on the role of depositors of Bank A and receive information on Bank A's fundamentals. Specifically, all depositors of a Bank A receive a signal of the form:

Bank A has [type] fundamentals.

This statement is correct with probability p.

Type describes the quality of Bank A's fundamentals (i.e., Good, Medium, or Weak). Each group of three depositors that form a Bank A is shown only one of these potential values. Systematically varying the value of *p* across disclosure treatment conditions allows us to effectively implement three levels of disclosure for Bank A: (1) No-disclosure, in which the signal is non-informative (p = 33%), meaning that it is equally likely for Bank A to be Good, Medium, or Weak; (2) Partial-disclosure, in which the signal is partially-informative (p = 66%) and reveals the most likely type;[48] and (3) Full-disclosure, in which the signal is fully-informative (p = 100%) and does not leave any room for uncertainty about Bank A's fundamentals. It is common knowledge that all members of a depositor group receive the same signal about their respective Bank A and decide simultaneously on whether to withdraw or not.

Interbank linkages disclosure

The second dimension we are interested in concerns the linkages (in form of assets commonality) between Bank A and a second bank (Bank B, hereafter), for which there is no explicit disclosure. Each participant in the experiment is a depositor at both banks and plays once the three-person coordination game with each bank (i.e., first with Bank A, and then with Bank B).[49] Depositors receive the following information regarding their respective Bank B:

With probability q, Bank B has the same fundamentals as Bank A.
This statement is always correct.

We vary the value of *q* to implement two distinct levels of disclosure about interbank linkages between the two banks: (1) non-informative disclosure about linkages (q = 33%), in which the type of Bank B is completely independent of the type of Bank A since disclosure about Bank A fundamentals provides no information about fundamentals of Bank B; and (2) partial disclosure (q = 66%), in which the two banks share the same type of fundamentals in two thirds of the cases.[50] Participants know that all depositors in their respective Bank B have received the same linkage information. It is also common

[48] If the actual bank type does not match the type signaled, both of the remaining types are equally likely. This is made explicit on the decision screens. Implementation of the disclosure treatments is explained in Appendix 5B.

[49] This design is consistent with evidence on banks' customers preference for maintaining multiple banking relationships and can be rationalized by assuming that depositors in Bank B have already some prior information about Bank A's fundamentals before receiving additional information about the potential linkages between these two banks.

[50] As for the type signals for Bank A, if the types of the two banks do not match, the other types are equally likely. Implementation of the linkage treatments is explained in Appendix 5B.

knowledge that their fellow Bank B depositors have received the same signal about Bank A, both with respect to the type of fundamentals and level of disclosure. At the time depositors take the withdrawal decision for Bank B, the uncertainty about the fate of Bank A (i.e., how many depositors have withdrawn and whether the bank has failed or not) has not yet been resolved. However, Bank B depositors are reminded about the specific type of signal they received for Bank A on the decision screen.

Our treatments allow us to simultaneously study the behavioral effects of different types of information on depositor behavior, as well as potentially resulting contagion effects from Bank A to Bank B in a unified setting. To this end, we systematically vary the degree of disclosure about Bank A's fundamentals and about the linkages between Bank A and Bank B (i.e., the degree to which information about the financial health of Bank A is relevant for assessing the health of Bank B) in different treatment groups.

5.4.3 Procedures and Supplementary Data

The experiment was programmed and conducted using z-Tree (Fischbacher 2007). A total of 432 participants were recruited using both hroot (Bock et al. 2014) and ORSEE (Greiner 2015).[51] One half of the experimental sessions were conducted at AWI Lab in Heidelberg, the other half at mLab in Mannheim. We conducted 24 sessions with 18 participants taking part in each session.[52] Each session was structured as follows: First, participants were given general information about the session and the payoff modalities. They learned that they would be paid for two parts of the experiment and receive further instructions at the beginning of each task. Participants proceeded to part 1, the bank run game. They were first given the instructions on screen and received a paper handout summarizing bank payoffs. Participants were asked to answer comprehension questions on the instructions and could only continue with the experiment after correctly answering all of them. They received feedback on the correctness of their answers, were given the opportunity to refer back to the instructions, and could correct their answers. They could also ask for assistance from the experimenters, although hardly anyone did. After the

[51] Two participants requested their data to be deleted after the experiment, leaving us with data from 430 participants. In two sessions data from the final demographics questionnaire was not correctly saved to disk. A total of 18 questionnaires could be restored from z-Tree Gamesafe files. No behavioral data was lost.

[52] The dataset as well as the complete script of the experiment are available online at https://ckgk.de/files/thesis/ch5_bank_runs.zip

comprehension questions, participants subsequently took the withdrawal decisions for Banks A and B on two separate screens.

For the purpose of the bank run game, each participant was randomized into two separate groups of three players each. One group represented the depositors of Bank A; the other one represented those of Bank B. Our protocol made sure that the group composition always differed between Bank A and Bank B in at least one participant.[53] Participants were matched in a way that also ensured that all depositors of the same Bank, i.e. members of a group, received identical information about their two Banks. Both coordination games, i.e. the one for Bank A as well as the one for Bank A, were payoff relevant.

To get insights into the channels through which bank disclosure affects behavior in the bank run game in the absence (or presence) of interbank linkages, we also elicited participants' beliefs. For both banks, participants were asked to indicate their beliefs about how many of the other depositors (i.e., none, one, two) they thought would choose to withdraw and how confident (0 – 100%) they were in this judgement. We ask for confidence to get an individual level estimate for the perceived strategic uncertainty in the decision situation. For Bank B, we additionally asked participants to indicate their beliefs about how likely (0 – 100%) it was for Bank B to be of the type indicated by the signal about Bank A.[54] To be least obtrusive, yet as close to participants' thought processes as possible, the unincentivized belief elicitation questions appeared on the same screens and at the same time as the payoff-relevant withdrawal decisions.

In part two, we also assess participants' attitudes towards losses. Loss aversion has been reported to affect behavior in coordination games (see Cachon and Camerer 1996, Rydval and Ortmann 2005, Trautmann and Vlahu, 2013). We implement Gächter et al.'s (2010) incentivized lottery choice task to elicit individual loss attitudes. The loss attitude elicitation followed immediately after the withdrawal decisions for the two banks. Participants received their payment for the loss aversion task in addition to the payoffs from the bank-run game in part one.

[53] Appendix 5B shows group assignments for both bank types.

[54] Note that we refrain from eliciting this belief for Bank A, because it would only present a trivial sanity check and not actually provide us with meaningful additional information about the way depositors approach the decision at hand.

Finally, at the end of each session, we collected demographics (age, gender, field of studies) and information on banking habits (number of bank accounts, customer of multiple banks, owning a savings account). Our participants are on average 22.6 years old, 52.4% are female, and 27.9% study economics. In terms of banking relationships, participants on average have 2.2 bank accounts with 64.4% owning a savings account. 57.0% of our participants hold accounts at more than one bank.

Participants' payment consisted of a show-up fee, payoffs for the bank-run game, and the payoff for the loss aversion task. On average, participants earned EUR 8 and the sessions lasted approximately 40 minutes.

5.5 Hypotheses

Hypothesis 1 (*Individual bank disclosure*). Conditional on the underlying bank type (i.e., Good, Medium, Weak), increased precision of disclosure about Bank A's fundamentals reduces the propensity of deposit withdrawal for banks with strong fundamentals (i.e., Good and Medium banks). Conversely, increased precision of disclosure about Bank A's fundamentals increases the propensity of bank withdrawal for banks with poor fundamentals (i.e., Weak banks).

This prediction derives from the literature reporting differential effects of financial information disclosure depending on the economic context (Bouvard et al. 2015, Leitner 2014, Nier 2005). Thus, we conjecture that reducing the uncertainty about a bank's type from full uncertainty (as is the case of No-disclosure treatment, when p = 33%) to none (as is the case of Full-disclosure treatment, when p = 100%), leads to more coordination and is beneficial for Good and Medium banks, but aggravates the coordination problem for Weak banks.

The following channel may be at work here: When disclosure reduces the uncertainty about a bank's type, it also affects the beliefs about the other bank's depositors' behavior. For those banks with strong fundamentals, more precise information about a bank's strength may increase the belief that the other depositors will keep the money in the bank. This in turn will reduce the propensity of withdrawing. The reverse holds for the banks with poor fundamentals.

Hypothesis 2 (*Absence of interbank linkages*). When the disclosure about interbank linkages is non-informative, the withdrawal decisions of Bank B's depositors are independent of their information about Bank A's type and the precision of that information.

This prediction derives from the fact that the type of Bank B is completely independent of the type of Bank A. In this framework, the disclosure about Bank A's fundamentals does not provide any information about the fundamentals of Bank B. Thus, we conjecture that Bank B's withdrawal rates will not exhibit significant variation conditional on the signal about Bank A's type and the precision of that signal.

Hypothesis 3 (*Partial interbank linkages*). When the disclosure about interbank linkages is informative, the withdrawal decisions of Bank B's depositors are correlated with the withdrawal decisions of Bank A's depositors across banks' types. The correlation is stronger for higher precision of disclosure about Bank A's type.

This prediction derives from the fact that in the presence of (partial) interbank linkages, disclosure about the types of Bank A provides a (noisy) signal about the type of Bank B. As a result, depositors in Bank B can learn about their bank's type from the disclosure about Bank A. Thus, we conjecture that on the one hand, when the signal about Bank A's type is non-informative, the withdrawals rates from Bank B will not exhibit significant variation across different signals about Bank A's type. On the other hand, as the precision about Bank's A type increases, the strength of bank fundamentals leads to more coordination towards repayment for Good banks than for Medium and Weak banks.

The following channel may be at work here: When disclosure about Bank A's type is non-informative, it has no effect on the beliefs about Bank B's type or on the beliefs about the behavior of other Bank B depositors. Thus, the pattern for withdrawals across banks' type is similar with that for Bank A in absence of disclosure. However, as the disclosure about Bank A's type becomes more precise, it affects the beliefs of Bank B's depositors about their bank's type, as well as the beliefs about other depositors' behavior. When more precise information about Bank A's type reveal that Bank A has strong fundamentals, information about partial linkages between Bank A and Bank B increases the belief that Bank B also has strong fundamentals while reducing the belief that the other Bank B depositors withdraw their money. These changes in beliefs in turn reduce the propensity of withdrawing.

5.6 Results

5.6.1 Overview

We first analyze withdrawal behavior from the two banks, focusing on the effects of disclosing information about Bank A. Then we consider Bank B and its economic linkages to Bank A. Here, we study how varying degrees of disclosure about one bank affects the withdrawal decision from the second in the presence (or absence) of economic linkages between the two. Having analyzed behavior on the level of individual withdrawal decisions, we move to expected bank failure rates to gain a better understanding of the consequences of depositors' actions. Finally, we dig deeper into the potential transmission channels, i.e. we try to uncover *how* our treatments lead to changes in behavior.

5.6.2 Individual Bank Disclosure

Table 5.2 presents the withdrawal behavior from Bank A, contingent on bank type and on the precision of disclosure about bank's type. We find statistically significant differences in withdrawal rates across the three disclosure levels and for all bank type signals. As shown in the first column of Table 5.2, for depositors who receive the type signal Good, the percentage of withdrawals drops significantly from 12.5% to 0% when the signal is partially informative rather than non-informative. Under Full-disclosure, the withdrawal rate is 2.1%, which is not statistically significantly different from the withdrawal rate in the Partial-disclosure condition (p = 0.32), but remains statistically significantly different from the No-disclosure treatment. These withdrawals rates suggest that Good banks benefit from increased disclosure.

Table 5.2: Withdrawals from Bank A

Bank A Type signal	Good			Medium			Weak	
No-disclosure	12.5%			31.3%			25.0%	
		**						**
Partial-disclosure	0.0%		**	33.3%		*	47.9%	
					**			
Full-disclosure	2.1%			15.2%			39.6%	

Notes: The table shows the percentage of participants in each condition that chose to withdraw. The brackets signify two-sided tests of proportions. */**/*** denote statistical significance at 10%/5%/1%; N = 46-48 in each group.

From the second column we observe that when depositors receive the signal Medium, the withdrawal rate does not differ significantly between No-disclosure and Partial-disclosure treatments ($p = 0.83$). However, the difference in withdrawal rates between Partial-disclosure and Full-disclosure is statistically significant, while the difference between No-disclosure and Full-disclosure is marginally statistically significant. These findings suggest that Medium banks benefit from full disclosure only, since partial information does not seem to be enough to significantly affect depositors' behavior.

Finally, for banks with a Weak type signal we observe a statistically significant increase in withdrawals between No-disclosure and Partial-disclosure treatments. The difference in withdrawal rates between Partial- and Full-disclosure, as well as that between No-disclosure and Full-disclosure, remain statistically insignificant ($p = 0.42$ and $p = 0.13$). In contrast to the other bank types, more precise disclosure is detrimental for Weak banks, which are more likely to suffer from liquidity problems triggered by reduced uncertainty about their assets' quality.

We speculate that the biggest difference in terms of information for depositors might actually be the switch from having no information at all to having at least some information, irrespective of it being partially or fully informative. Thus, we pool the data from both disclosure treatments and compare it to the No-disclosure condition. The results are reported in Table 5.3. We observe that disclosure of any kind significantly reduces withdrawals from banks with a Good type signal and significantly increases withdrawals from banks with Weak type signal compared to the No-disclosure conditions. For depositors who receive a Medium type signal, the differences in withdrawal rates are not significantly different between the No-disclosure and Disclosure conditions ($p = 0.39$). The results for aggregated disclosure conditions are generally in line with those based on the fully differentiated treatment conditions and sharpen the picture: Disclosure works to reduce withdrawals from banks which are believed to have strong fundamentals, but aggravates the situation for those believed to have weak fundamentals. The results for Bank A are generally consistent with hypothesis 1.

Table 5.3: Withdrawals from Bank A with Pooled Disclosure Conditions

Bank A Type signal	Good		Medium	Weak	
No-disclosure	12.5%	***	31.3%	25.0%	**
Disclosure	0.0%		24.5%	43.8%	

Notes: The table shows the percentage of participants in each condition that chose to withdraw. The brackets signify two-sided tests of proportions. */**/*** denote statistical significance at 10%/5%/1%. N = 48 for No-disclosure, N = 94-96 for Disclosure.

Table 5.4: Withdrawals from Bank B (No-linkages)

Bank A Type signal	Good	Medium	Weak
No-disclosure	33.3%	20.8%	29.2%
Disclosure	31.3%	27.1%	18.8%

Notes: The table shows the percentage of participants in each condition that chose to withdraw. N = 24 per group in No-disclosure, N = 48 per group in Disclosure.

5.6.3 Interbank Linkages Disclosure

Next, we analyze the behavior of depositors in Bank B. This allows us to identify the impact of disclosure about Bank A's type on their withdrawal decisions, both in absence and presence of interbank linkages between the two banks. First, we focus on the No-linkages condition, for which all depositors know that the probability of both banks having the same type is 33%. The columns in Table 5.4 show depositors' withdrawal rates from Bank B contingent on different signals about Bank A's type. Having realized in the previous section that the distinction between partial and full disclosure is of minor importance to depositors, we pool both treatments for the analysis of withdrawals from Bank B[55]. Neither in the No-disclosure nor in the Disclosure setting we are able to find any statistically significant differences in pairwise proportions testing of the withdrawal rates from Bank B across Bank A's type (comparing along the rows, within the two disclosure conditions). At the same time, we also do not find any statistically significant differences in the withdrawal rates from Bank B across disclosure conditions, holding the signal about Bank A constant (i.e. comparing along the columns). In the absence of interbank linkages between the two banks, depositors do not seem to (inadequately)

[55] Appendix 5C shows the results without pooling the data.

transfer information disclosed about Bank A to Bank B, i.e. we do not find any evidence for financial contagion in the absence of interbank linkages. This result is in line with our hypothesis 2.

Table 5.5: Withdrawals from Bank B (Partial-linkages)

Bank A Type Signal	Good	Medium	Weak
No-disclosure	16.7%	29.2%	29.2%
Disclosure	12.5%	26.1%	33.3%
	Good vs. Medium: *		Good vs. Weak: **

Notes: The table shows the percentage of participants in each condition that chose to withdraw. The brackets signify two-sided tests of proportions. */** denote statistical significance at 10%/5%. N = 24 per group in No Disclosure, N = 46-48 per group in Disclosure.

Next, we report the results from the Partial-linkages condition, for which withdrawal rates from Bank B are depicted in Table 5.5. We first compare withdrawal rates along the rows. If depositors know that there is a two-thirds probability for Bank B having the same type as Bank A, but they do not have any information about the type of the latter (No-disclosure), withdrawal rates from Banks B do not differ statistically significantly across the three type of signals. In contrast, if depositors do receive valuable information about Bank A, they also take the presence of interbank linkages between the two banks into account when making their withdrawal decision for Bank B. In the presence of interbank linkages and meaningful disclosure about Bank A, the withdrawal rates from Bank B are statistically significantly lower if the signal for Bank A is Good rather than Weak. The difference in withdrawals from Bank B when the signal about Bank A's type reveal Good rather than Medium fundamentals remains marginally statistically significant. However, there is no statistically significant difference in withdrawals from Bank B between Medium and Weak type signals ($p = 0.45$). These observations are consistent with our third hypothesis, i.e. information disclosed about Bank A is only used for Bank B, if the information is meaningful for this institution.

Again, it is also possible to compare withdrawal rates from Bank B in the Partial-Linkages condition along the columns. That is, we can hold the type signal for Bank A constant and compare withdrawal rates from Bank B between No-disclosure and

Disclosure conditions. None of the pairwise t-tests reveals statistically significant differences in withdrawal rates (all p-values > 0.6).

Finally, we directly compare the No-linkages and Partial-linkages treatments. That is, we compare withdrawal rates from Banks B between the two linkage conditions, holding the type signal received for Bank A as well as the disclosure regime constant. Table 5.6 shows the results. We find that only in the Disclosure condition and in the presence of a Good type signal about the fundamentals of Bank A, there is a statistically significant difference between the withdrawal rates from Bank B between No-linkage and Partial-linkage conditions. The fact that the difference is not significant for Weak type signals despite the fact that the withdrawal rates under Partial-linkages is almost twice as high as in the case of No-linkages can most likely attributed to insignificant power (18.8% vs. 33.3%, two-sided test of proportions, p = 0.104). After all, the disclosure for Bank A only provides a very noisy signal for the type of Bank B, given that the types are equal in only two thirds of all cases.

Table 5.6: Withdrawals from Bank B by Linkage Condition

Bank A Type signal	Good		Medium		Weak	
Condition	No-linkage	Partial-linkage	No-linkage	Partial-linkage	No-linkage	Partial-linkage
No-disclosure	33.3%	16.7%	20.8%	29.2%	29.2%	29.2%
Disclosure	31.3%	12.5%	28.1%	26.1%	18.8%	33.3%
	**					

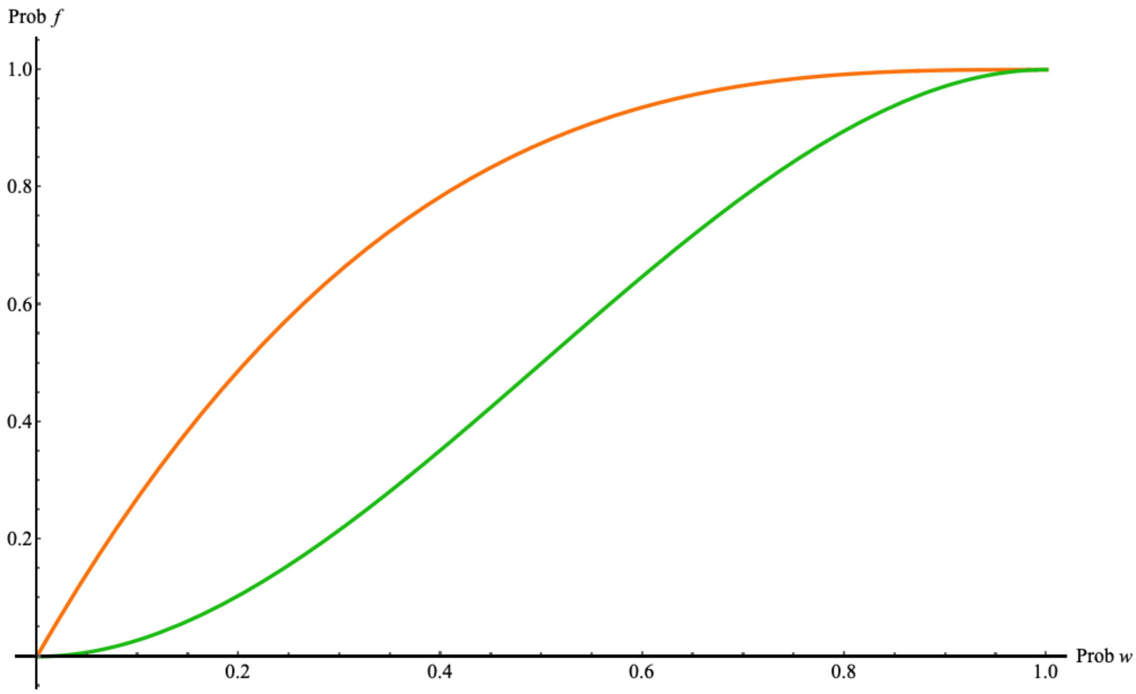
Notes: The table shows the percentage of withdrawals in each condition. The brackets signify two-sided tests of proportions. ** denotes statistical significance at 5%. N = 24 per group in No-disclosure, N = 48 per group in Disclosure.

Our results show that the information about potential linkages between financial institutions may affect the impact of disclosure about individual banks on depositors' behavior. If depositors at one bank are aware of such linkages between their bank and another bank in the economy, they correctly process and transfer the disclosed information. We furthermore observe a statistically significant "flip effect" of disclosure for Good and Weak banks B.[56] That is, Good banks tend to benefit from greater precision

[56] Fisher's exact test of withdrawals from Bank B in Disclosure condition, categories: No-linkage / Partial-linkage and Good / Weak type signal; p = 0.02.

in disclosure, while Weak banks tend to suffer and are more likely to end up failing. This result for withdrawal rates from Bank B is in line with our observations for Bank A, for which we also found that meaningful disclosure affects Good banks positively and Weak banks negatively. Despite the fact that the type signal for Bank A only provides very noisy information about the type of Bank B in the Partial-linkages conditions, the behavioral pattern appears to be very similar.

Figure 5.1: Bank Failure Probabilities



Notes: Prob *f* denotes the probability of bank failure. Prob *w* denotes the probability of withdrawal. Graph for Medium and Weak banks in orange (upper); graph for Good bank type in green (lower).

5.6.4 Bank Failures

Apart from looking at individuals' withdrawal behavior, we can also examine expected outcomes of the bank run coordination games.[57] The probability of a bank failure to occur depends directly on the probability that a randomly selected depositor withdraws. In turn, the probability of withdrawal is affected by the information a depositor has about their banks. In our setup, banks of Good type fail if two or more depositors withdraw. Banks of Medium or Weak type fail if at least one depositor withdraws. Thus, depositors'

[57] We consider expected coordination outcomes rather than the actual outcomes in our experiments, because our total number of banks is relatively low and coordination outcomes depend on the depositor composition of each bank. As an example, consider 9 depositors in 3 banks of Weak type. If 3 of the 9 depositors withdraw, we could observe anywhere from one to three bank failures, depending on how depositors are randomized into the depositor groups.

withdrawal propensities translate into expected bank failures differently.[58] Figure 5.1 shows the relationship between the withdrawal probabilities and the probability of bank failure for the three types. Table 5.7 summarizes the probability of bank failure for our various treatment conditions and presents them side-by-side with observed withdrawal rates.

Table 5.7: Probability of Bank Failure

Type Signal	Good		Medium		Weak	
	w	f	w	f	w	f
Bank A						
No-disclosure	12.5%	4.3%	31.3%	67.6%	25.0%	57.8%
Partial-disclosure	0.0%	0.0%	33.3%	70.3%	47.9%	85.9%
Full-disclosure	2.1%	0.1%	15.2%	39.0%	39.6%	78.0%
Bank B (No-linkages)						
No-disclosure	33.3%	25.9%	20.8%	50.3%	29.2%	64.5%
Disclosure	31.3%	23.3%	27.1%	61.3%	18.8%	46.5%
Bank B (Partial-linkages)						
No-disclosure	16.7%	7.4%	29.2%	64.5%	29.2%	64.5%
Disclosure	12.5%	4.3%	26.1%	59.6%	33.3%	70.3%

Notes: Columns w show withdrawal rates; columns f show the corresponding expected bank failure probabilities. These are calculated by treating observed withdrawal rates as withdrawal probabilities.

Bank failure probabilities help us to understand the effects different withdrawal rates have for the variuos bank types. For example, if one third of the depositors of Good banks withdraw, this only leads to a probability of bank failure of 25.9%. In contrast, for Medium and Weak types in our setup, the same withdrawal probability translates into a 70.3% probability of bank failure (approx. 2.7 times as high). While individual depositors' withdrawal behavior might not be of biggest interest to policy makers and regulators, bank failures clearly are. This is because of the large number of depositors affected as well as the ripple effects bank failures can produce in the financial system. The exercise of calculating bank failure probabilities from observed withdrawal decisions highlights how small changes in depositor behavior interact with the potentially

[58] For Good types the proability of bank failure F_G depending on withdrawal probability p is given by $F_G(p) = 3p^2 - 2p^3$. For Medium and Weak types it is $F_{M,W}(p) = 1 - (1-p)^3$.

unobservable fragility of financial institutions to produce large differences in economic outcomes.

5.6.5 Transmission Channels

Having studied actual withdrawal behavior and observed large differences in the probabilities for observing subsequent bank failure, we now look at the mechanisms underlying the behavioral effects. As hypothesized, differences in withdrawal behavior in response to our treatment conditions could be resulting from changes in the beliefs of depositors about the type of their banks as well as the behavior of their fellow depositors. Different precision levels of the type disclosure for Bank A directly inform participants about the likelihood of encountering each type of bank. This should affect their belief about how many of the other depositors, who have received the same information, withdraw their money.

First, we need to establish that beliefs individuals have about the number of other depositors withdrawing their money from the bank correlate with actual withdrawal decisions. We asked participants to indicate how many other depositors they think would withdraw their money from Bank A. We find a strong, positive, and statistically highly significant correlation between individuals belief about how many of the others would withdraw and their actual withdrawal decision (Spearman's rho = 0.71, $p < 0.01$). There is also a strong correlation between the believed number of other withdrawals and participants' own withdrawal decision (Spearman's rho = 0.55, $p < 0.01$) for Bank B. While the correlation is slightly less pronounced than for Bank A, it still points to widespread consistency between beliefs and actions. This holds for the No-linkages as well as the Partial-linkages conditions (rho = 0.56, $p < 0.01$ and rho = 0.54, $p < 0.01$). That is, higher numbers of believed withdrawals are associated with a higher propensity to withdraw. Participants rationally react to the expected behavior of their fellow depositors. The next step is to assess how our treatment variations affect the beliefs that participants form about the two banks.

Bank A

We observe a positive and statistically highly significant correlation between the type signal about Bank A (coded as 1 = Good, 2 = Medium, 3 = Weak) and the believed number of withdrawals (withdrawals are 0, 1, 2, Spearman's rho = 0.29, $p < 0.01$). That is, signals of lower bank quality are associated with a higher number of expected withdrawals. Depositors appear to correctly form their beliefs about others behavior based on the bank type signal.

Depositors also take disclosure (type signal precision) into account when forming their beliefs about the behavior of others based on the signals they receive: In the No-disclosure treatment, in which the type signal is not informative, the correlation between signal type and believed number of withdrawals is low and only marginally statistically significant (rho = 0.1621, p = 0.052).[59] The correlation is much stronger and highly statistically significant in both treatments in which the signal is at least partially informative (Partial-disclosure: rho = 0.32, p < 0.01; Full-disclosure: rho = 0.40, p < 0.01). As expected, more precise type signals affect beliefs more strongly. The better the information available to depositors, the more they differentiate between the types.

Instead of holding the disclosure level constant and analyzing the correlation between signal type and believed number of withdrawals, we can also hold the signal constant and look at the correlations between signal precision and the believed number of withdrawals. For Good type signals, we find the correlation to be negative and highly statistically significant (rho = -0.26, p < 0.01). That is, the more informative the type signal for banks believed to be Good, the fewer depositors are expected to withdraw. For Medium signals, the correlation is lower and of reduced statistical significance, but still negative (rho = -0.2, p = 0.02). For Weak signals, finally, the correlation is essentially zero (rho = 0.04, p = 0.6). While increases in the precision of disclosed information seems to matter for the belief formation for Good and Medium type signals, it is not associated with any significant differences in beliefs in case of Weak quality signals.

While the previous analysis hints at interaction effects between disclosure and signal types, it does not allow us to adequately assess these. We therefore turn to a multivariate regression framework, which also allows us to include additional control variables. Initially, we regress the believed number of withdrawals on the level of disclosure, the bank type signal, and their interaction by means of an ordered probit regression.[60] In a later step, we also add controls for age, gender, loss aversion, being an economist, owning a savings account, having multiple bank accounts, banking with multiple banks, and

[59] For some reason, participants still seem to react to the different words used in the instructions (Good / Medium / Weak).

[60] OLS regressions yield qualitatively similar results. The ordered probit model better fits the discrete dependent variable.

having participated in Mannheim rather than in Heidelberg. The estimation results are shown in Table 5.8.

Table 5.8: Multivariate Analysis of Withdrawal Beliefs for Bank A

	(1)	(2)
Partial-disclosure	-0.220	-0.104
	(0.265)	(0.273)
Full-disclosure	-1.010***	-0.869**
	(0.338)	(0.353)
Medium signal	0.337	0.432*
	(0.248)	(0.263)
Weak signal	0.494**	0.565**
	(0.247)	(0.257)
Partial-disclosure x Medium signal	0.354	0.157
	(0.355)	(0.374)
Partial-disclosure x Weak signal	0.549	0.442
	(0.353)	(0.368)
Full-disclosure x Medium signal	0.373	0.033
	(0.430)	(0.465)
Full-disclosure x Weak signal	1.178***	1.065**
	(0.413)	(0.433)
Age		0.003
		(0.015)
Economist		-0.161
		(0.159)
Female		0.250*
		(0.141)
Mannheim		0.069
		(0.135)
# of Bank Accounts		-0.0389
		(0.056)
Multiple Banks		0.209
		(0.151)
Savings Account		-0.048
		(0.145)
Loss Aversion		-0.068
		(0.064)
Observations	430	398

Notes: Ordered probit model. Standard errors in parentheses. Dependent variable: Belief about how many other depositors in the group will withdraw. Base categories: No-disclosure and Good type signal. */**/*** denote statistical significance at 10%/5%/1%.

The regression results reveal that the main driving factors behind the beliefs about the number of withdrawals are Full-disclosure (reduces withdrawal beliefs), receiving a Weak signal (increases withdrawal beliefs), and the combination of both situations (increases withdrawal beliefs strongly). These results are robust to the addition of control variables. Our data paints a clear picture for Bank A. Different levels of disclosure interact with the signal about a bank's type to influence depositors' belief about how many depositors will withdraw their money from the bank. Furthermore, beliefs translate into actual withdrawal decisions in a way that can be described as individually rational. It appears that one channel through which disclosure of bank stability information affects withdrawal behavior is through a change in beliefs about other depositors' likely actions. This observation is in line with our first hypothesis.

Bank B

The picture changes if we turn towards Bank B. Beliefs about the number of withdrawals do not correlate statistically significantly with either the signal about Bank A or the level of disclosure. While this is expected in the absence of interbank linkages, it is quite surprising in their presence.

We also probe these observations in a multivariate framework to uncover potential interaction effects of type signal and the level of disclosure. The model specifications follow those of Bank A. We estimate the models with and without our set of controls as well as separately for the case of No-linkages and Partial-linkages. Table 5.9 shows the ordered probit regression results.[61] Apart from the coefficient of having accounts with multiple banks, no other covariates reach statistical significance. Confirming our initial observations from simple correlations, the belief about the number of other depositors withdrawing from the bank is not significantly affected by the level of disclosure, the signal about Bank A or their interaction in either linkage condition.

[61] We estimate the models separately for the two conditions to avoid the inclusion of a triple interaction term, which is notoriously hard to interpret. OLS regressions yield qualitatively similar results.

Table 5.9: Multivariate Analysis of Withdrawal Beliefs for Bank B

	(1) No-linkages	(2) No-linkages	(3) Partial-linkages	(4) Partial-linkages
Partial-disclosure	0.192	0.344	0.372	0.492
	(0.323)	(0.333)	(0.336)	(0.356)
Full-disclosure	-0.069	-0.007	-0.070	0.004
	(0.327)	(0.338)	(0.346)	(0.379)
Medium Signal	-0.150	-0.051	0.226	0.289
	(0.330)	(0.340)	(0.339)	(0.372)
Weak Signal	-0.357	-0.233	0.138	0.216
	(0.334)	(0.344)	(0.343)	(0.358)
Partial-disclosure x Medium	0.0855	0.032	-0.105	-0.419
	(0.461)	(0.482)	(0.471)	(0.512)
Partial-disclosure x Weak	-0.275	-0.316	0.0570	0.023
	(0.473)	(0.487)	(0.472)	(0.498)
Full-disclosure x Medium	0.342	0.228	-0.596	-0.695
	(0.465)	(0.487)	(0.507)	(0.565)
Full-disclosure x Weak	0.419	0.303	0.148	-0.137
	(0.468)	(0.481)	(0.484)	(0.544)
Age		-0.009		-0.030
		(0.019)		(0.030)
Economist		0.046		-0.225
		(0.211)		(0.217)
Female		0.161		0.245
		(0.182)		(0.193)
Mannheim		0.108		0.252
		(0.178)		(0.183)
# of Bank Accounts		0.006		-0.154
		(0.037)		(0.099)
Multiple Banks		0.354*		0.457**
		(0.183)		(0.224)
Savings Account		0.1000		0.248
		(0.187)		(0.219)
Loss Aversion		0.098		0.027
		(0.078)		(0.082)
Observations	216	207	214	191

Notes: Ordered probit model. Standard errors in parentheses. Dependent variable: Belief about how many other depositors in the group will withdraw. Base categories: No-disclosure and Good type signal. */**/*** denote statistical significance at 10%/5%/1%.

In stark contrast to the results for Bank A, we do not find a statistically significant influence of our treatments on the beliefs participants form about the number of other depositors withdrawing from Bank B. As established before, beliefs still translate into choices, but it is less clear how beliefs are formed for Bank B in the first place. Given that the link between banks A and B is partial at best, the signal participants receive about the type of Bank B is also very noisy. Taken together with the reduced number of observations for the two linkage conditions, we might be running into power issues that prevent us from clearly identifying the causal factors underlying belief formation.

5.7. Conclusion

In this article, we study the fundamental mechanism of information disclosure about the fragility of financial institutions. In line with our hypotheses and the literature, we find that the effects of increased precision in the information disclosed depend on the financial institutions' fundamentals. If banks are believed to have strong fundamentals and thus a large capacity to withstand liquidity shocks, disclosure that is more precise serves to reduce the likelihood of bank runs by reducing the probability of customers withdrawing their deposits before asset maturity. In contrast, banks believed to have weak fundamentals face larger likelihoods of early withdrawals if the signal about the banks fundamentals becomes more precise. Our belief data suggests that disclosed information affects beliefs about the number of depositors that is expected to withdraw. Participants then react accordingly.

In addition, our results suggest that disclosing meaningful information at all compared to not releasing any information, has larger effects on depositors' withdrawal decisions than different levels of precision in the disclosed information. Both of these observations speak directly to the policy question of publicly releasing bank stress test results discussed in the introduction. They underline the need for regulators to take into account the differential effects disclosure can have for banks with solid or fragile fundamentals and carefully weigh the potential for positive and negative consequences of publishing stress test results.

Another aspect we study is the transmission of information disclosed about one financial institution to another. Notably, we are able to study both information disclosure and the transmission of information in a unified setting, which can be easily adapted to incorporate different levels of disclosure precision and linkage levels. We distinguish two cases: One in which there are no economic linkages between financial institutions and another one in which interbank linkages exist in the form of a positive probability for similar asset exposure. In the absence of interbank linkages, if information disclosed about one bank were to systematically affect depositors' likelihood of withdrawing their money from another bank, information would be inadequately applied to an unrelated entity. We do not find any evidence for this problematic form of financial contagion in our experiment.

In the second setting, in which we model partial interbank linkages, we observe that information disclosed about one bank also affects withdrawal rates for the linked institution. In this case, the disclosed information about one bank also provides a meaningful, yet noisy, signal about the fundamentals of the second bank. In our experiment, depositors are able to identify that the information is valuable for both institutions and act accordingly. In this regard, we find support for the findings by Brown et al. (2017). Our result can be interpreted in the context of the debate about the publication of stress test results. As mentioned before, stress tests typically only cover a subsample of financial institutions in an economy. Yet, test results published for these banks can also affect other banks which were not covered in the stress test. This can be the case if institutions are believed to be similar to the tested ones in terms of business models, asset exposures, or other forms of interbank linkages that enable the disclosed information to be perceived as meaningful.

Appendix 5A

Table 5A.1: Calibration of Payoffs in the Experiment

Bank Type and Own Decision	Number of Other Depositors Withdrawing		
	0	1	2
Good			
not withdraw	210	210	0
withdraw	85	45	30
Medium			
not withdraw	210	0	0
withdraw	60	30	20
Weak			
not withdraw	150	0	0
withdraw	60	30	20

Notes: All payoffs are in experimental currency units.

Table 5A.1 shows payoffs to participants in the bank-run coordination games in experimental currency units. Survival of a bank always yields a higher payoff to its depositors than failure. Depositors keeping their money in a surviving bank are not influenced by the other depositors' decisions and thus earn the highest payoff. For Good and Medium banks, the high payoffs are identical. Assets of Weak banks are assumed to earn a lower return than those of Good and Medium ones. Consequently, for Weak banks the high payoff is lower than that of the other two.

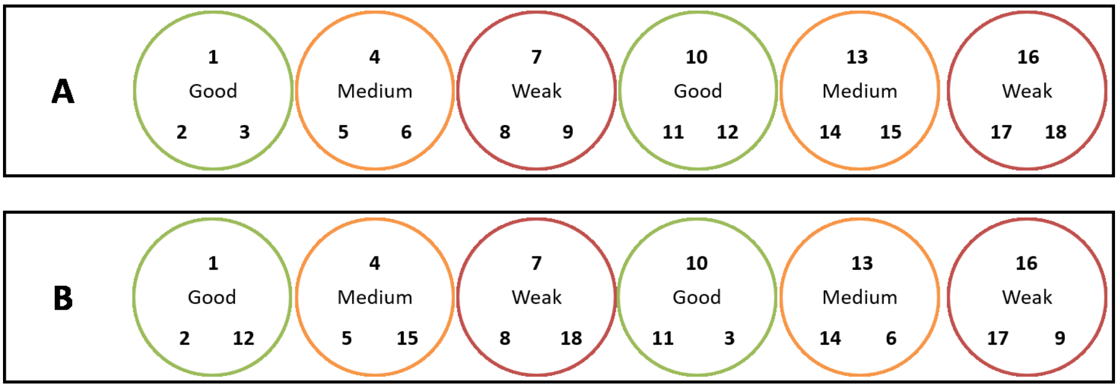
If only a single depositor withdraws from a Good bank, it survives and is not liquidated. However, the withdrawing depositor forgoes some future return. The resulting payoff may not be larger than the liquidation value of the bank, because otherwise, the bank would have to be liquidated to cover the claim.

Banks also differ in fragility. While Good banks fail conditional on at least two of its depositors withdrawing, Medium and Weak banks already fail if at least one of the depositors withdraws. If a bank fails, depositors who have decided not to withdraw their money lose everything and are left with a payoff of 0. Depositors withdrawing from a failing bank divide the available funds (i.e., the liquidation value) between themselves. Naturally, if fewer people withdraw the individual shares are larger. The payoffs correspond to the available funds (i.e., the liquidation value) divided by the number of depositors withdrawing. Good banks have a higher liquidation value than Medium and Weak banks. The differentiating factor of Medium and Weak banks is their ability to earn returns on customer’s deposits which is assumed to be lower for Weak banks than for Medium ones.

Appendix 5B

In each session, there were 18 participants. Participants were randomly assigned cubicles in the laboratory. The cubicles were always matched in the same way to ensure an equal number of banks of each type in all sessions and treatment conditions. There always were two Banks A of each type (Good, Medium, Weak) and two Banks B of each type in each session. Figure 5B.1 shows how cubicles, numbered from 1 to 18, were matched to bank types.

Figure 5B.1: Cubicle to actual bank type matching



Notes: The first row only shows Banks A, the second shows Banks B. Each circle represents a bank, i.e. a depositor group in the coordination game. Green (Orange, Red) circles represent Good (Medium, Weak) type banks. Depositors are represented by cubicle numbers.

Example: The first Bank A of type Good consists of participants sitting at cubicles 1 to 3. For depositors at cubicles 1 and 2, their Bank B is also of Good type. In Bank B, their third depositor is the participant in cubicle 12. Their fellow Bank A depositor in cubicle 3, however, is part of the fourth Bank B, which is of also of Good type. For each participant, banks A and banks B never consist of the same set of depositors.

Note that the figure shows the actual bank types, which participants typically do not know for sure. The only case in which they can be certain of a bank's type occurs in the Full Disclosure treatment, in which they know their Bank A's type for sure. The way we implement group matching allows us to make truthful statements about the probabilities of banks A and B having the same types in our linkages treatment, while at the same time ensuring that we can implement all information disclosure precision levels for banks of type A.

Table 5B.1 shows the Bank A type signal each individual receives. It depends on a random draw, which is automatically conducted by the computer at the beginning of each session. This random draw determines which of the different sub-cases of each treatment condition is implemented. Each case (within a treatment condition) is equally likely. The random draws ensure that the probabilities of each signal being correct are truthful. Take the No-Disclosure treatment as an example. Each one of the three cases is implemented with 1/3 probability. Depending on the case, the members of exactly one bank type (Good, Medium, or Weak) receive a signal that perfectly corresponds (in its type) to the actual bank's type. As participants are randomized to player numbers (cubicles in the lab), there is a chance of exactly 1/3 that their bank actually has the type given by the signal. A similar argument holds for the Partial-Disclosure treatment. In 2 out of 3 cases, participants receive a signal that matches their actual type of bank.

Table 5B.2 shows Bank B types for each participant. Again, a computerized random draw at the beginning of the session determines which of the cases is implemented. Note that the cases in this treatment directly determine the actual type of Bank B for each participant, rather than a signal about its type. This is the result of participants receiving a statement about the probability that their Bank B is of the same type as Bank A. In the No-Linkages treatment and in each of its cases, the members of exactly one type of Bank A (Good, Medium or Weak) face a Bank B which is of the same type as A. In the Partial-linkages treatment and in each of its cases, the depositors of two Bank A types face a Bank B which is of the same type as A.

Table 5B.1: Bank A – Types and Signals

Player	Bank A	Bank B	Type A	Signal A: No-Disclosure *c1*	*c2*	*c3*	Partial-Disclosure *c1*	*c2*	*c3*	*c4*	*c5*	*c6*	Full-Disclosure
1 2 3	1	1 4	G	G	W	M	G	G	G	G	W	M	G
4 5 6	2	2 5	M	W	M	G	M	M	M	M	G	W	M
7 8 9	3	3 6	W	M	G	W	W	W	W	W	M	G	W
10 11 12	4	4 1	G	G	W	M	G	G	G	G	W	M	G
13 14 15	5	5 2	M	W	M	G	M	M	M	M	G	W	M
16 17 18	6	6 3	W	M	G	W	W	W	W	W	M	G	W

Notes: G/M/W denote Good/Medium/Weak bank type. c1 to c6 denote cases 1 to 6.

Table 5B.2: Bank B – Types

Player	Bank A	Bank B	Type A	Type B: No-Linkage *c1*	*c2*	*c3*	Partial-Linkage *c1*	*c2*	*c3*	*c4*	*c5*	*c6*
1 2 3	1	1 4	G	G	W	M	G	G	G	G	W	M
4 5 6	2	2 5	M	W	M	G	M	M	M	M	G	W
7 8 9	3	3 6	W	M	G	W	W	W	W	W	M	G
10 11 12	4	4 1	G	G	W	M	G	G	G	G	W	M
13 14 15	5	5 2	M	W	M	G	M	M	M	M	G	W
16 17 18	6	6 3	W	M	G	W	W	W	W	W	M	G

Notes: G/M/W denote Good/Medium/Weak bank type. c1 to c6 denote cases 1 to 6.

Appendix 5C

Table 5C.1: Withdrawals from Bank B (No-linkages)

Bank A Type signal	Good	Medium	Weak
No-disclosure	33.3%	20.8%	29.2%
Partial Disclosure	25%	37.5%	12.5%
		**	
Full Disclosure	37.5%	16.7%	25%

Notes: The table shows the percentage of participants in each condition that chose to withdraw. ** denotes statistical significance at 5%. N = 24 per group.

Table 5C.2: Withdrawals from Bank B (Partial-linkages)

Bank A Type signal	Good	Medium	Weak
No-disclosure	16.7%	29.2%	29.2%
Partial Disclosure	20.8%	33.3%	33.3%
Full Disclosure	4.2%	18.2%	33.3%
		***	

Notes: The table shows the percentage of participants in each condition that chose to withdraw. *** denotes statistical significance at 1%. N = 24 per group.

Chapter 6

Countercyclical Risk Aversion: Beyond Financial Professionals

Abstract. We test if Cohn et al.'s (2015) experimental results on countercyclical risk aversion exhibited by financial professionals generalize to a standard student sample. In our sample, we do not find an effect of stock market bust or boom on subjects' investments. We do not find a systematic emotional reaction, nor do we find an effect of variation in the emotional state (especially fear) on investment. Our results add to the literature documenting behavioral differences between financial professionals and non-professionals and, taking a policy perspective, underline the need for careful external validity checks of single sample experiments.[62]

[62] This chapter is co-authored by Stefan Trautmann and published as König-Kersting and Trautmann (2018).

6.1 Introduction

An ever-growing literature documents behavioral differences between financial professionals and non-professionals. Some authors find financial professionals to be less affected by behavioral biases (e.g. Alevy et al. 2007, Feng and Seasholes 2005, Kaustia et al. 2008, Kirchler et al. 2018, List 2003; Shapira and Venezia 2001). Others report them to be more affected than non-professionals (Haigh and List 2005, Gilad and Kliger 2008). These findings have two consequences: First, generalizability of effects found for any specific subgroup to the population cannot safely be assumed. Second, if the results of these studies are supposed to adequately inform policy makers and regulators, careful identification of the relevant sample for policy purposes is warranted. Are financial markets predominantly driven by the behavior of professionals with substantial investment amounts, or are they driven by the behavior of a large population of retail investors through their choice of mutual funds, pension products, and mortgages?

In this paper, we revisit the experimental evidence for countercyclical risk aversion by Cohn et al. (2015, hereafter CEFM). In their experiment, financial professionals are primed with either a stock market boom or bust and subsequently take incentive compatible investment decisions over risky assets. CEFM observe that participants who are primed with a bust scenario invest significantly less than those primed with a boom scenario – that is, the participants exhibit countercyclical risk aversion. The authors argue that rendering the bust scenario mentally salient increases fear, which in turn affects participants' propensity to take risks.

We ask whether countercyclical risk aversion is a general phenomenon or whether it might be restricted to financial professionals through either self-selection or learning. Thus, we probe the generalizability of CEFM's findings by transferring their experimental setup from a sample of financial professionals to a standard student sample. We are interested in seeing whether the strong emotional and behavioral reactions to the stock market priming of financial professionals carry over to the student population (economics and non-economics). In the future, they will belong to the high-income segment of the general population, and will likely be retail investors; indeed, some of our participants do already invest regularly in the stock market. At the same time, they are a more heterogeneous group and less affected by corporate culture in the financial industry than the financial professionals in CEFM. If the effects observed by CEFM transfer to this population, countercyclical risk aversion and the fear-channel may have implications for the broader set of financial decisions made by typical retail investors.

We implement the original experiment as closely as possible and only alter design elements, which, in their original form, proved unsuitable to employ in the (German) student sample. Our sample size (N = 200) allows to identify the main behavioral treatment effect on investment in the risky asset reported by CEFM with a power ≥ 0.9.[63] Moreover, we also test for differences in investments in the ambiguous asset, thereby increasing the probability to detect an existing effect.

In our experiment, the treatment effect sizes of the bust vs. boom priming on investment decisions fall short of statistical significance. Moreover, the student sample does not show a systematic emotional reaction to the stock market priming as measured by self-reported general affective state and fear. Moreover, we do not find a significant effect of emotional state on investment decisions. Thus, in contrast to the financial professionals in CEFM, the students in our experiment do not exhibit countercyclical risk aversion.

The structure of this paper closely follows CEFM for easy comparison. We first present the experimental design and the laboratory procedures. We then present the results with respect to investment decisions, emotional reaction, and the effect of emotions on investment. A discussion of the findings concludes the paper. For convenience, figure and table numbers throughout this paper correspond to those of the original study in a one-to-one fashion.

6.2 Experimental Design

6.2.1 Design

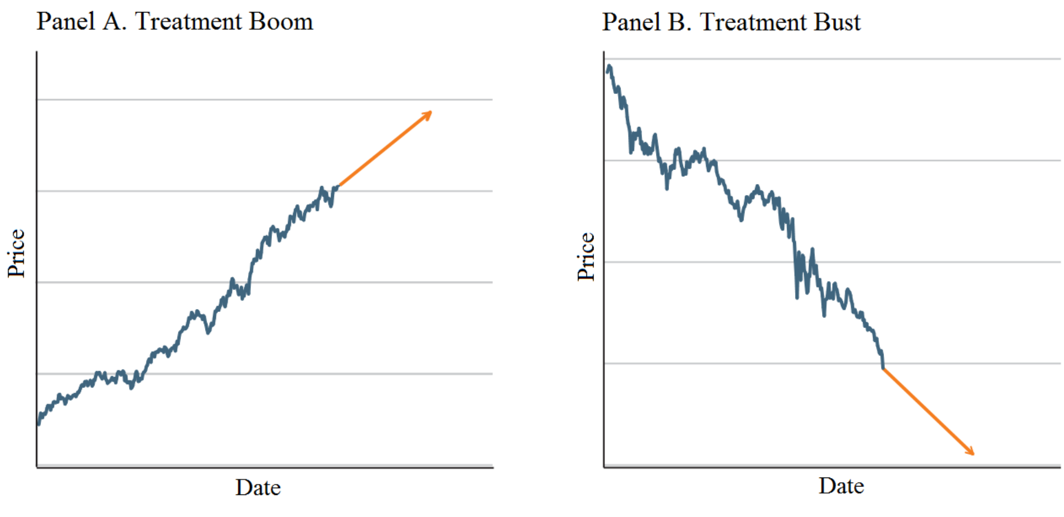
As the original experiment, our experiment begins with what CEFM call icebreaker questions regarding participants' trading behavior. These questions cover self-reported trading frequency, investment behavior, trust in financial advisors as well as patience

[63] CEFM, second investment decision (risky asset with known probabilities), average investment shares: Boom = 57.71 (s.d. 29.25), Bust = 45.20 (s.d. 30.26), difference 12.51. Total sample size required to detect an effect of this magnitude with same standard deviations with power 0.9 in a one-sided t-test of independent groups: $N_1 = 196$ (two-sided $N_2 = 240$).

with regard to financial decisions.[64] We have added a question from the German Socio-Economic Panel (SOEP) asking participants to indicate their readiness to take financial risks on an 11-point Likert scale. The measure is included to observe whether there is an effect of the priming with respect to the responses in the incentivized behavioral measure of risk attitudes later in the experiment, controlling for the reported baseline risk attitude.

The next part of the experiment is the priming phase, which is identical to the original study. Depending on treatment, participants are either presented with an animated graph of a continuing stock market boom or bust and are asked to explain whether they would invest in stocks, precious metals, exchange traded funds, real estate or hold cash given the presented market dynamics. CEFM kindly provided the original graphics of their study. We reprint their Figure 6.1 for illustrative purposes below. In the experiment, each of the priming questions is presented on a separate screen and no time or word limits are enforced. Afterwards, participants report their general affective state using a 9-point version of the self-assessment manikins (Bradley and Lang 1994, Irtel 2008) and current level of fear on a 7-point Likert scale (Bosman and van Winden 2002).

Figure 6.1: Investment Decisions by Task and Treatment



Notes: This is a close adaptation of Figure 1 from CEFM. It shows the charts used in the priming parts of both the original and our study. The graphs are animated and reveal the price data from left to right. The orange arrows indicate that the trends are expected to continue for the foreseeable future.

Subsequently, participants take two investment decisions, one of which would become payoff relevant for one fifth of the participants of each session. While the

[64] The Online Appendix includes additional figures, tables, the glossary of financial terms and the complete script of the experiment. The Online Appendix and the Replication Package, which includes our data and data analysis files, are available at https://doi.org/10.11588/data/HLGUFF.

financial professionals in CEFM were endowed with CHF 200 each, our participants were endowed with EUR 25. Adjusting endowments to income levels is common practice in experiments involving different population samples as it serves to ensure setting appropriately adjusted incentives (Alevy et al. 2007; Haigh and List 2005). While our students' endowments are lower in absolute terms than those of the professionals, the stakes are quite sizable for this population.[65] In both decision tasks, participants are asked how much of their endowment they would like to invest in a risky asset, keeping the remainder for sure. The tasks include a verbal description as well as a photo of the physical box, which represents the asset. Each box is filled with a combination of yellow, red and blue balls.[66] After both investment decisions and all questionnaires have been completed by all participants, one ball is drawn from each box. If it is yellow, participants win 2.5 times their invested amount. If it is blue or red, participants lose their investment.[67] In the first investment decision, the asset is characterized by ambiguity in the sense that the exact winning probability is unknown. The corresponding box is filled with a large number of balls in all three colors. Participants are asked to guess the share of yellow (winning) balls, which provides an individual measure of expectation. In the second task, the winning probability is known to be 50%: the box contains exactly one yellow (win) and one red (lose) ball.[68]

Following the investment tasks, participants complete a questionnaire on their perspective on life, including the general optimism question used by the original authors and taken from the standard Life Orientation Test (Scheier and Carver, 1985; Scheier et al. 1994). Due to the different population studied, several adjustments were necessary in this part of the experiment. We exchanged the Swiss Market Index (SMI) for the German Market Index (DAX) in the question on market outlook to increase familiarity. We also replaced the questions on the likelihood of losing the current job with a (arguably more optimistic) question regarding the expectation of finding a job after completing their studies. In addition, we dropped the question on participants' perception of income

[65] CEFM report an average monthly gross income of over CHF 11,000. The student sample indicates median net income to be EUR 250-500 per month. Based on these numbers, the endowment constitutes about 1.8% of the monthly income for the professionals and 6.7% for the students in our study.

[66] As in the original experiment, the physical boxes were visibly placed in the front of the laboratory.

[67] Participants could win up to EUR 62.5 from the investment task. To prevent participants from leaving the laboratory empty handedly, everyone received an additional show-up fee of EUR 3.

[68] All participants take the two tasks in the same order. Just as in the original study, the tasks are not counterbalanced.

relative to the population average. Given that students are typically not employed full-time but are still in education, responses to questions regarding losing the current job or their relative income cannot be expected to be comparable to the original study. In order to keep the length and structure of the questionnaire as close to the original as possible, we still included questions designed to capture similar aspects of life.

The second to last questionnaire is the financial literacy test. CEFM designed their own quiz, because commonly used ones appeared to be too easy for financial professionals. With similar reasoning, we use a financial literacy test that better fits the general population. We use four questions covering interest rates, inflation as well as the consequences of stock and funds ownership from van Rooij et al. (2011) as well as Lusardi and Mitchell (2011). As in the original study, participants are asked to self-assess their performance on the financial literacy test in terms of the number of questions answered correctly. The final part of the survey is a short demographic survey, closely following the original study. We exchanged the question for the professional function with a question on the field of studies, adapted the monthly income question for Germany and removed the question on liquid wealth. Most students probably belong to the lowest category of liquid wealth ownership, effectively providing no variation that could be exploited in the data analysis. Furthermore, removing questions about personal wealth from the questionnaire reduces the perceived intrusiveness of the experiment. As in CEFM, participants could leave comments and remarks on the study.

6.2.2 Procedure

The experiment was implemented on SoSciSurvey.de, an online questionnaire platform by SoSci Survey GmbH, which is free to use for academic purposes. It was run at the experimental laboratory of the Alfred-Weber-Institute for Economics at the University of Heidelberg. Participants were recruited using both ORSEE (Greiner 2015) and hroot (Bock et al. 2014). After arriving at the laboratory and registering for the session, respectively half of the participants were randomly assigned to the boom and bust treatments by the computer. They were not aware of the random assignment to different treatments and could individually complete the experiment at their own pace. In contrast to the original experiment, participants were provided a short glossary of finance terms used throughout the experiment to facilitate better understanding.[69] One paper glossary was placed in each cubicle for easy reference. Throughout the session, the two boxes containing the balls stood visibly on a table in the front of the room. After all participants had finished the computerized part of the experiment, the boxes were covered up so that their contents were hidden. Seat numbers were drawn from an urn (without replacement) to determine the participants whose investment decision were to be implemented for real.[70] For each of these participants, a second random draw determined which of the two investment decisions became payoff relevant, and one ball was drawn from the corresponding physical box to determine payoffs. The color of the ball drawn and the amount invested in the relevant investment task determined payments. Participants received their payments in cash and left the lab.

6.3 Results

6.3.1 Summary Statistics

In total, 200 students participated in the experiment over the course of 12 sessions. Their average age at the time of the experiment was 23.3 years, 45% are male and 32% are economists.[71] Participants reported a median net income level of EUR 250-500 per month. Compared to CEFM, we have a lower share of males, our participants are generally younger, and earn less. All of these differences are natural consequences of our choice to conduct the experiment with a standard student sample. With gender being controlled for in the multivariate analysis, we decided against enforcing the gender split of 75% males reported in the study of financial professionals. On average, our participants answered three of the four financial literacy questions correctly, which, in contrast to the original study, is not significantly correlated with their trading frequency. Concerning the latter, our student sample is quite active on financial markets: 60% reported to be trading assets at least once a year and 15.5% even indicated to be trading securities at least once per month.

---

[69] The glossary is available in the Online Appendix. It contains brief explanations for the German equivalents of the following financial terms: investment vehicle, Exchange Traded Fund (ETF), DAX, classic investment fund, and risky financial investment.

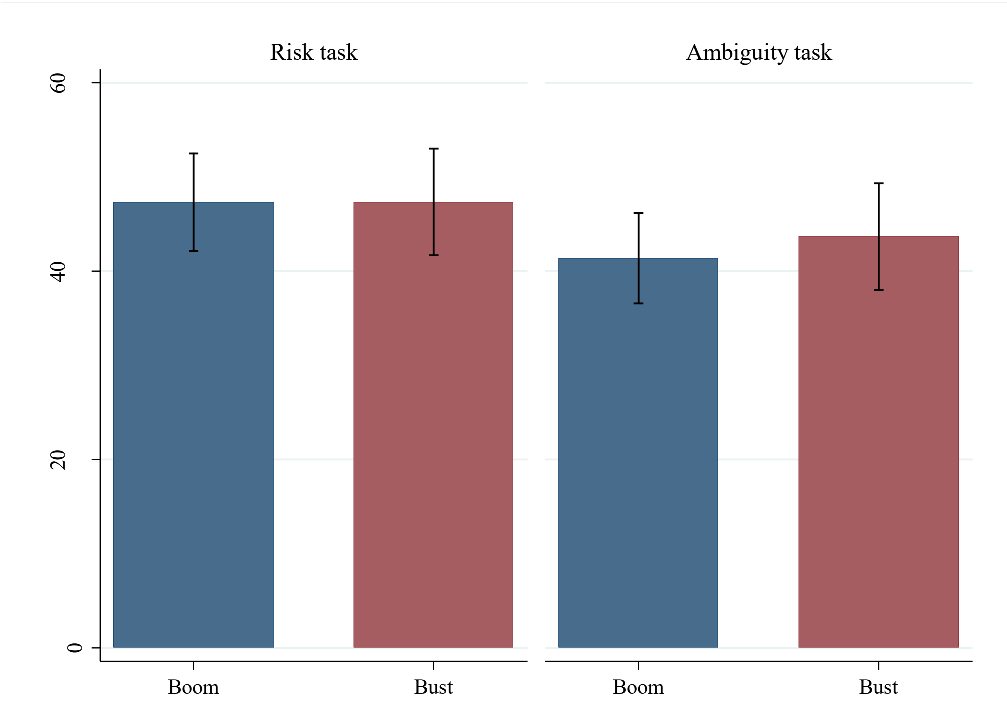
[70] To determine the number of decisions to be implemented for real, we divided the number of participants present by five and rounded up to the next integer. This ensured that at least 20% of the participants were selected for payment in each session. In the end, 44 of 200 participants had their choices implemented.

[71] With 32%, economics is the most prevalent field of study in our sample. It is followed by law (9%), political science (8%), and sociology (6.5%). A complete breakdown is available in the Replication Package.

6.3.2 Investment Decisions

Figure 6.2 shows average investment shares in the boom and bust conditions separately for the risk and ambiguity tasks. Average investments in the bust condition are higher than in the boom condition, by 0.1% in the risk task and 5.6% in the ambiguity task. Neither difference is statistically significant (two-sided t-tests, N = 200; risk: p = 0.99; ambiguity: p = 0.53). In contrast, CEFM find reduced risk taking in the bust condition of 22% and 17% respectively for the risk and the ambiguity tasks. Notably, the picture does not change if we restrict the hypothesis testing to subgroups of our sample which are closer to financial professionals in terms of knowledge of finance and personal experience with financial markets: There are no significant differences in investment shares between treatment conditions for students of economics, participants scoring highest (4/4) in the financial literacy quiz, participants being active on financial markets at least once a month, and those old enough to be at least of legal age during 2010's Eurozone crisis.[72] We conclude that the stock market priming does not seem to affect the risk taking of non-professionals in the proposed direction in our experiment.

Figure 6.2: Investment Decisions by Task and Treatment



Notes: The figure shows average investment shares with error bars.

[72] All tests are two-sided t-tests. Economists only: N = 63; risk: p = 0.18; ambiguity: p = 0.64. Financial literacy: N = 93; risk: p = 0.85; ambiguity: p = 0.64. High trading frequency: N = 31; risk: p = 0.78; ambiguity: p = 0.68. Age: N = 55; risk: p = 0.44; ambiguity: p = 0.63.

As in the original study, we probe these initial observations in a multivariate regression framework to include control variables from the accompanying questionnaires. We estimate the following models:

$$(1)\quad y_{ik} = \beta_0 + \beta_1 Bust_i + \beta_2 Ambiguity + \beta_3 \boldsymbol{X}_i + \varepsilon_{ik}$$

$$(2)\quad y_{ik} = \beta_0 + \beta_1 Bust_i + \beta_2 Ambiguity + \beta_3 Bust_i \times Ambiguity + \beta_4 \boldsymbol{X}_i + \varepsilon_{ik}$$

The dependent variable y_{ik} denotes the percentage share of the endowment the individual i invested in the asset in task k. $Bust_i$ and *Ambiguity* are indicators for decisions in the ambiguity task and the bust condition. $\boldsymbol{X}_i$ is a set containing the control variables age, gender, financial literacy, and trading frequency. Finally, ε_{ik} is the error term of the OLS regression with standard errors clustered at the individual level. Regression results are reported in Table 6.1, columns one and two.

Table 6.1: Regression Analysis of Investment Decisions

	Share invested in risky asset	
	(1)	(2)
Bust	1.113	-0.030
	(3.475)	(3.823)
Bust x Ambiguity		2.287
		(2.903)
Ambiguity	-4.840***	-5.961**
	(1.462)	(2.376)
Age	0.104	0.104
	(0.317)	(0.317)
Male	14.230***	14.230***
	(3.659)	(3.664)
Financial literacy	-3.185*	-3.185*
	(1.913)	(1.915)
High trading frequency	-4.491	-4.491
	(4.718)	(4.724)
Constant	48.899***	49.460***
	(9.583)	(9.662)
Observations	400	400

Notes: We report OLS coefficient estimates with standard errors clustered on the individual level in parentheses. The dependent variable is the share invested in the asset. Bust and Ambiguity are indicators for treatment and task. High trading frequency is an indicator for individuals who trade securities at least once per month. ***/**/* indicate significance at 1%/5%/10%.

In model (1), similar to CEFM, participants invest a significantly lower share of their endowment in the ambiguous asset compared to the risky asset. That is, ambiguity aversion seems to be an aspect of the current setting that is very robust with respect to the variation in the population studied. In contrast to the original paper, the most important coefficient estimate for the Bust treatment is positive, yet it does not reach statistical significance. When including the interaction of treatment and investment task indicators in model (2), the coefficient on the Bust indicator is basically zero. The interaction itself also remains insignificantly different from zero. Independent of the specification, we find a strong and robust difference in the invested share between men and women, with the former investing significantly more in the asset than the latter.[73] The same tendency is found in the original paper, albeit non-significantly so. We observe a marginally significant, negative correlation between the score in the financial literacy quiz and the invested amount. This stands in contrast to the coefficient estimates by CEFM, where the effect is (insignificantly) positive. However, given the different questions as well as the different sample, the contrasting effects are not surprising and can potentially be explained by a multitude of factors.

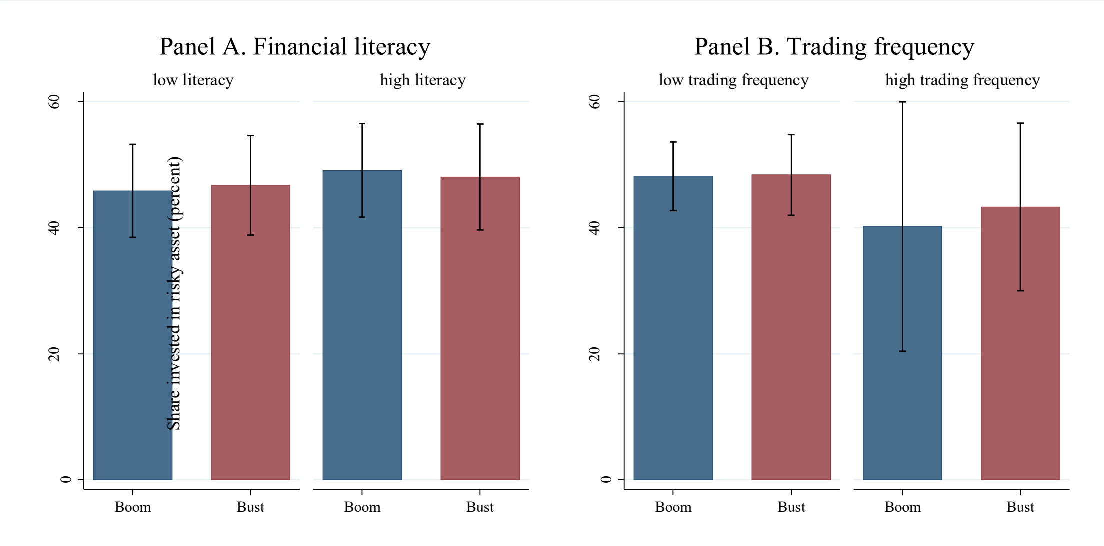
Following the original paper, we treat financial knowledge and trading frequency as proxies for participants' market experience and separately compare average investment shares of participants with high and low financial literacy respectively trading frequencies. For both variables, we employ a median-split rule to create indicator variables. Figure 6.3 visualizes the data. The figure for our sample does not show a pattern of investment decisions similar to the original study with financial professionals. CEFM reported a very clear pattern: Average Boom investment shares were always significantly higher than the corresponding Bust investment shares. In contrast, we find investment shares in the risk task to be basically the same across all conditions.[74] OLS regressions which follow model (1) but include interaction terms of bust and financial literacy, respectively trading frequency, show that there are no significant differences in the reaction to the priming treatment depending on financial literacy or trading frequency.[75]

[73] We also run separate regressions for male and female participants based on models (1) and (2), as well as a regression based on model (1) which includes an interaction term between the Bust and Male indicators. In all cases the results remain qualitatively the same. The regressions are available in the Replication Package.

[74] For the ambiguity task, see Figure A1 in the Online Appendix.

[75] These results are available in Table A1 in the Online Appendix.

Figure 6.3: The Role of Market Experience



Notes: The figure shows average investment shares in the risk task with error bars.

Table 6.2: Regression Analysis of Expectations

	Guessed probability of success	General Optimism
	(1)	(2)
Bust	-1.987	-0.114
	(1.876)	(0.193)
Age	-0.459***	0.015
	(0.148)	(0.021)
Male	1.624	-0.319
	(2.004)	(0.198)
Financial literacy	0.355	0.155
	(1.173)	(0.113)
High trading freq.	2.722	0.077
	(2.119)	(0.275)
Constant	55.880***	4.226***
	(5.077)	(0.601)
Observations	200	200

Notes: We report OLS coefficient estimates with standard errors clustered on the individual level in parentheses. The dependent variable in column one is the guessed share of winning balls in the ambiguity task. In column two, it is self-reported general optimism. Bust indicates the bust priming condition. High trading frequency is an indicator for individuals who trade securities at least once per month. ***/**/* indicate significance at 1%/5%/10%.

6.3.3 Expectations

As part of the first investment decision, the ambiguity task, participants are asked to guess the share of yellow (winning) balls in the large box. Their answers provide an individual measure of expectation. In line with CEFM, we now test whether the stock market priming has an effect on expectations. To do so, we run an OLS regression of participants' guessed share of winning balls in the ambiguity task on a treatment indicator and the usual set of control variables. In addition, we also consider participants self-reported general optimism as a second dependent variable that could be affected by the priming treatment. The results are reported in Table 6.2, columns one and two. The regression results do not indicate a statistically significant effect of the priming condition on the guessed probability of success or participants' general optimism. In this regard, the results are identical to the original study.

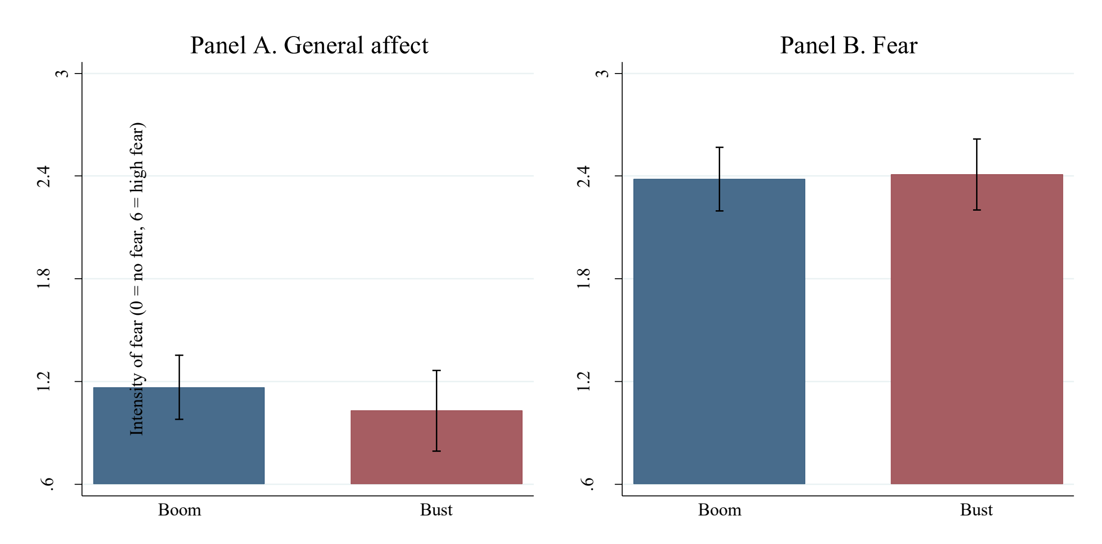
6.3.4 Emotions

One of the key findings of CEFM is the apparent channel through which their stock market priming affects risk taking behavior. In conjunction with their fear induction experiment[76], which we do not transfer to the current setting, they argue that rendering the stock market crash salient increases participants' fear which in turn reduces their willingness to take risks. As in the original study, we collect information on participants' general affective state as well as their self-reported level of fear. Figure 6.4 shows how these measures are affected by the two treatment conditions. In contrast to CEFM, average general affect scores and average fear scores are basically identical in bust and boom. There is no significant treatment effect in either aspect. The same holds if we again restrict the analysis to those students, who have the highest involvement and experience with financial markets (see section 6.3.2).

Similar to the previous analyses, we estimate OLS regression models that include the emotion measures as the dependent and bust treatment indicators as the explanatory variables while also accounting for the set of controls. The results are reported in columns one and two of Table 6.3. In both estimations, the coefficient of the treatment dummy is not significantly different from zero, providing no basis for the rejection of the null hypotheses. As expected, the effect of the bust treatment on the measure of general affect is very close to zero. The coefficient of the general affect score is negative, while the coefficient of the fear score is very slightly positive. This observation matches CEFM's result in direction, but lacks statistical significance.

[76] CEFM also report on a second, separate experiment. This additional experiment presents experimental evidence of the effects of fear on risk-taking behavior. The experiment involves a student sample and induces fear by means of electrical shocks. It is not directly connected to the priming experiment which we revisit here.

Figure 6.4: Booms, Busts, and Emotions



Notes: The figure shows average scores with error bars for general affect and fear in the boom and bust conditions.

Table 6.3: Regression Analysis of Emotions

	General affect	Fear	Share invested in risky asset		
	(1)	(2)	(3)	(4)	(5)
Bust	-0.157	0.024			1.126
	(0.218)	(0.202)			(3.483)
General affect			-0.021		
			(1.117)		
Fear				-0.529	-0.532
				(1.034)	(1.041)
Age	-0.013	-0.017	0.097	0.088	0.095
	(0.022)	(0.017)	(0.313)	(0.315)	(0.318)
Male	0.224	-0.272	14.283***	14.136***	14.085***
	(0.218)	(0.216)	(3.658)	(3.641)	(3.637)
Financial literacy	0.227*	-0.085	-3.220*	-3.270*	-3.230*
	(0.120)	(0.109)	(1.915)	(1.908)	(1.925)
High trading frequency	0.321	-0.098	-4.284	-4.339	-4.543
	(0.332)	(0.273)	(4.724)	(4.702)	(4.742)
Ambiguity			-4.840***	-4.840***	-4.840***
			(1.462)	(1.462)	(1.464)
Constant	0.597	3.193***	49.685***	51.373***	50.599***
	(0.643)	(0.540)	(8.935)	(9.841)	(10.354)
Observations	200	200	400	400	400

Notes: We report OLS coefficient estimates with robust standard errors in parentheses. In columns three to five the standard errors are clustered on the individual level. The dependent variable in column one is self-reported general affect. In column two, it is self-reported fear and in columns three to five it is the share invested in the risky. Ambiguity and Bust indicate task and treatment conditions. High trading frequency is an indicator for individuals who trade securities at least once per month. ***/**/* indicate significance at 1%/5%/10%.

CEFM move on to estimate the effect of emotions on risk taking behavior in terms of the share invested in the risky assets. Although we do not find a treatment effect on emotions, there is still considerable variation across subjects how the priming task affects their emotions. This allows us to identify the proposed fear-investment channel: While none of the participants indicated a complete lack of fear, a little more than 60% of participants indicate low scores of 1 and 2 in both treatments. Medium scores of 3 and 4 were reported by 29% in boom and 23% in bust. Finally, 10% and 12%, respectively, reported even higher levels of fear. We thus also run these regressions and present the results in columns three to five of Table 6.3. In our sample, neither general affect, nor fear alone has a significant effect on the investment share (columns three and four). Consequently, including the fear score in addition to a bust indicator also does not result in the appearance of any significant effects in terms of the main variables of interest (column five).

6.4 Conclusion

We set out to test whether countercyclical risk aversion is a more general phenomenon than CEFM's experimental results suggest. Specifically, we ask whether a student sample shows the same behavioral reactions to the stock market priming as the financial professionals of the original study. The results of our experiment, which closely follows the original study despite some necessary minor adjustments, are very clear. CEFM's results do not extrapolate from financial professionals to the student sample, as none of the treatment effects reaches statistical significance. This is despite the fact, that we consider both investment decisions (under risk and under ambiguity) and use a sample size big enough to detect an effect of the original size and direction with a power of 0.9 for the risky investment task.

It has been suggested that priming works by making previous experiences salient. Indeed, studies by Callen et al. (2014), Cordes et al. (2017), and Malmendier and Nagel (2011) support the idea that personal experiences affect risk preferences. In this spirit, the students in our sample might not react to the treatment manipulation, because they lack the exposure to financial markets and their fluctuations. We can only address this concern by carefully investigating specific subgroups of our sample which can be understood as rather well-informed retail investors, and therefore arguably share some of the financial market experience of financial professionals. We find that even those participants who are old enough to experience the financial crisis of 2010 as adults, are studying economics, show highest financial literacy, or have highest trading activity do not show countercyclical risk aversion in response to our treatments.[77] It appears that factors beyond emotional and professional experience play a role in differentiating financial professionals from well-informed retail investors. In support of this notion, Cordes and Dierkes (2017) demonstrate that personal experience is not always sufficient: For West Germans, who were brought up in the Federal Republic of Germany, they find that having experienced positive macroeconomic developments positively affects their likelihood of stock market participation. The authors do not find this effect for East Germans who were raised in the German Democratic Republic. They conclude that the effects of macroeconomic experiences may be mitigated by prevailing societal norms and values.

As CEFM demonstrate in their second, physiological experiment, fear affects risk taking in a student population. We do not contest this mechanism. Taken together with the self-reported emotional reactions by financial professionals in the first experiment as well as the differences in investment decisions, their findings suggest a fear channel underlying countercyclical risk aversion for financial professionals. We find that the stock market priming condition does not provoke an emotional reaction in students and also fails to directly affect investment behavior.[78] We also do not find fear effects on the investment decision in our sample. Thus, the countercyclical element in risk attitudes, working through fear effects, may be limited to financial professionals and may not extrapolate to other groups of the population. In contrast to the apparently very subtle priming effects, ambiguity effects and gender differences in investment robustly emerge in our study, extrapolating CEFM's results to the student sample.

Our findings add to the growing literature documenting behavioral differences between financial professionals and non-professionals. Most recently, Kirchler et al. (2018) highlight these differences in the context of tournament rankings and risk-taking

[77] We run the regressions presented in Table 6.1 with the respective subsamples of our participants. Results are provided in Tables B1 to B4 in the Online Appendix. We find no evidence for countercyclical risk aversion in any of the four subsample tests. While sample sizes (and thus power) in each test are lower, we still observe clear evidence for ambiguity attitudes and gender differences in these subsamples.

78 We also run the regressions presented in Table 6.3 with the previously mentioned subsamples of our participants. Results are reported in Tables B5 to B8 in the Online Appendix. There is no emotional reaction to the bust-boom priming, nor are there any fear effects on investment shares, in any of the four subsample tests. Again, while sample sizes (and thus power) in each test are lower, we still observe clear evidence for ambiguity attitudes and gender differences in these subsamples.

with very large samples of both financial professionals and students. They show that the effectiveness of certain incentive mechanisms in affecting behavior can depend strongly on the sample and context considered. In similar vein, Weitzel et al. (2018) demonstrate that asset markets with financial professionals are significantly more efficient and less prone to bubbles than asset markets with student traders. It remains an open question whether the origins of these differences are to be found in self-selection or in learning and socialization in the profession.

One should also consider that a large fraction of university students are set to become well-paid professionals, who will take risky investment decisions with respect to pensions, insurances, and wealth accumulation just a couple of years later. Clearly, today’s students represent a relevant section of tomorrow’s market participants, who are likely to be actively involved in investment decisions, even if they seek investment advice at a bank. Taking a policy perspective, it is not obvious whether a financial professional or a general retail investor sample is most relevant for any given regulatory issue. At the very least, our results highlight the importance of questioning (and possibly directly testing) the generalizability of treatment effects identified in experiments conducted exclusively with special subsamples of the general population.

Chapter 7

oTree Manager: Multi-User oTree Installations Made Easy

Abstract. oTree Manager is a software package designed to support the setup and management of multiple, production-ready oTree installations on the same server. Running in Docker containers, the individual instances run completely independent of each other while being less resource-intensive than traditional virtual machine setups. A convenient web interface provides both experimenters and laboratory managers with easy access to the most common actions and eliminates the need for command line interaction. Finally, oTree Manager comes with a novel *Lobby* feature, which makes laboratory experiments that use oTree's rooms feature more convenient to run.[79]

[79] This chapter is currently under review (round two) for the special issue on experimental software of the Journal of Behavioral and Experimental Finance.

7.1 Introduction

The oTree software package is increasingly used for conducting experiments both in the laboratory as well as online (Chen et al. 2016). A common problem for laboratory managers when preparing their laboratory for running experiments programmed using oTree is how to handle multiple experimenters. oTree itself does not come with built-in multi-user support.[80] As such, it assumes that only a single experimenter uses an installation at any given time. While experimenters can technically share a single installation, this can lead to a multitude of problems: Experimental sessions might be interrupted and data might be lost if one experimenter installs their experiments while the other one is running a session. Resetting the database to prepare for a newly added experiment might delete data the other experimenter has already collected, but not yet downloaded. Being defined installation-wide rather than per experiment, experimenters also cannot run experiments with different language-, currency-, or experimental currency settings simultaneously. In short, experimenters might unintendedly affect each other's experiments negatively.

The official oTree documentation provides an outline of how to better handle these situations. The proposed solution is to give each experimenter their own oTree installation. This can be done either by manually setting up an individual environment for each experimenter or by running individual Docker containers.[81] The first option involves manually creating virtual environments, configuring PostgreSQL and Redis databases, and handling port allocations, which is a tedious and error-prone effort, even for skilled laboratory managers. While coming with pre-configured databases and oTree installations, the second option still requires manual configuration for each additional experimenter. By default, both pathways yield inconvenient URLs, which require experimenters to address the correct ports to connect to their oTree installation. Clearly, these solutions quickly become very time-consuming, especially if oTree instances need

[80] There is a glossary of technical terms at the end of the chapter. It also provides links to the software packages mentioned throughout the article to avoid cluttering the footnotes.

[81] The community resources also include downloadable virtual machine templates and management scripts created by Felix Albrecht and Holger Gerhardt (https://otree-virtual-machine-manager.readthedocs.io).

to be added or removed on a regular basis, for example, when the number of experimentalists using oTree is growing or turnover is high.[82]

This paper presents *oTree Manager*, which is designed specifically to address these issues by automating the necessary steps to give each experimenter their own oTree instance.[83] oTree Manager gives laboratory managers and experimenters an intuitive graphical user interface, which allows them to set up and manage multiple, independent oTree instances on the same server without involving any manual configuration. The user interface is web-based and built on the same strong foundations as oTree itself. It is accessible by both laboratory managers and experimenters. Laboratory managers can create oTree installations and assign them to experimenters. These, in turn, can log in to easily set the oTree web interface password, reset the databases, or restart their oTree instance.

The development of oTree Manager adheres to three main objectives: 1) The final software should be easy to use for lab managers and experimenters alike. This includes that users should have to interact with the command line as little as possible. 2) It should be reliable and as fault tolerant as possible. In this regard, it is especially important that instances cannot (negatively) affect each other.[84] 3) The software should be easy to update and maintain and should not require extensive programming knowledge to setup. The main goal is to make the lives of lab managers and experimenters easier. For a live demonstration, please visit https://demo.otree-manager.com.

Adhering to these principals comes at a cost. Dokku, a core software dependency of oTree Manager, only runs on Unix operating systems such as Debian Linux.[85] Windows and macOS are not supported. Because of this limitation, oTree Manager can only be installed on UNIX operating systems as well. At first sight, this fact seems to limit its usefulness for experimental laboratories relying on Windows machines, e.g. for historical reasons such as z-Tree compatibility. These systems, especially if they run on consumer hardware, are typically not designed for continuous, ‘24/7’ operation and are unsuitable for running oTree Manager as an always accessible web service. In contrast to server

---

[82] A common scenario are students who need individual instances to deploy their own experiments for course or thesis work.

[83] I use installation and instance synonymously.

[84] It is especially important that errors in or outright crashes of one instance must not affect other instances.

[85] It is suggested to run oTree Manager on Debian 9, for which installation scripts are provided as part of the online documentation at http://docs.otree-manager.com.

hardware and professional workstations with high CPU core counts and large amounts of system memory, these computers are also typically not optimized for running many virtualized containers, which limits the number of oTree instances that can be run in parallel with adequate performance. Thus, it is highly recommended to run oTree Manager on a dedicated server or workstation, possibly in parallel to the existing infrastructure. In these kinds of setups, operating system selection is not an issue.

The remainder of the paper is structured as follows: Section 7.2 presents the features oTree Manager offers to experimenters and laboratory managers. Section 7.3 explains oTree Manager's architecture and touches on its dependencies. Section 7.4 concludes the paper.

7.2 Features

7.2.1 Experimenters

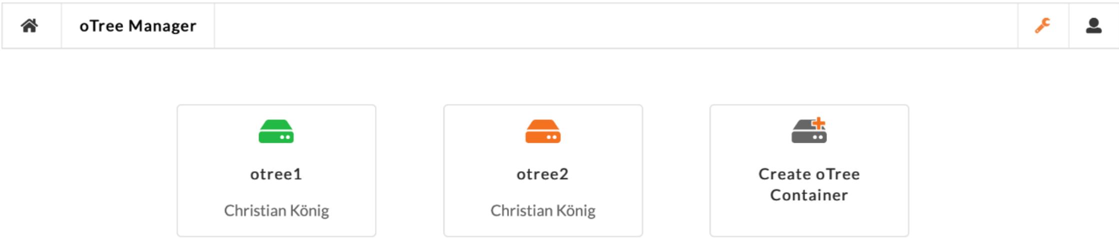
Experimenters can comfortably manage their oTree instance using an intuitive web interface (Figure 7.1), to which they receive individual user accounts. oTree Manager comes with full support for standard user credential management: Users can change their password or request a new one should they have forgotten their old one. They can also set and reset their deployment keys, which allow for secure, password-less transfers of the experiments to the server. In oTree Manager, oTree instances are always associated with exactly one experimenter account (but experimenters can have multiple oTree instances).[86] Experimenters can only see and configure their own instances, adhering to the principle of limiting access to only what is strictly necessary.

oTree instances set up by oTree Manager are pre-configured for immediate use. They come with production-grade databases (PostgreSQL and Redis), oTree's authorization features are enabled, strong admin passwords are set and debug mode is turned off. Experimenters do not need to configure these aspects manually. The instances come with Git access for easy deployment of experimenter's code. Experimenters can add the repository as a new remote (just as they would for a cloud-hosted server on

[86] It is important to distinguish different aspects: While only one experimenter account is linked to an oTree instance within oTree Manager, the experimenter is free to share the credentials to the standard oTree web interface with colleagues, for example to enable different experimenters to conduct sessions of the same experiment. However, there should only be one experimenter who can modify oTree server settings, reset the database or push new versions to the server. If there is more than one, miscoordination can adversely affect running sessions or lead to data loss. Thus, as a safety measure, each instance is linked to exactly one experimenter account within oTree Manager.

Heroku etc.) and simply push their latest version to the server. Instances managed by oTree Manager automatically detect newly pushed code revisions, install all required dependencies, and restart the oTree instance.

Figure 7.1: oTree Manager Dashboard



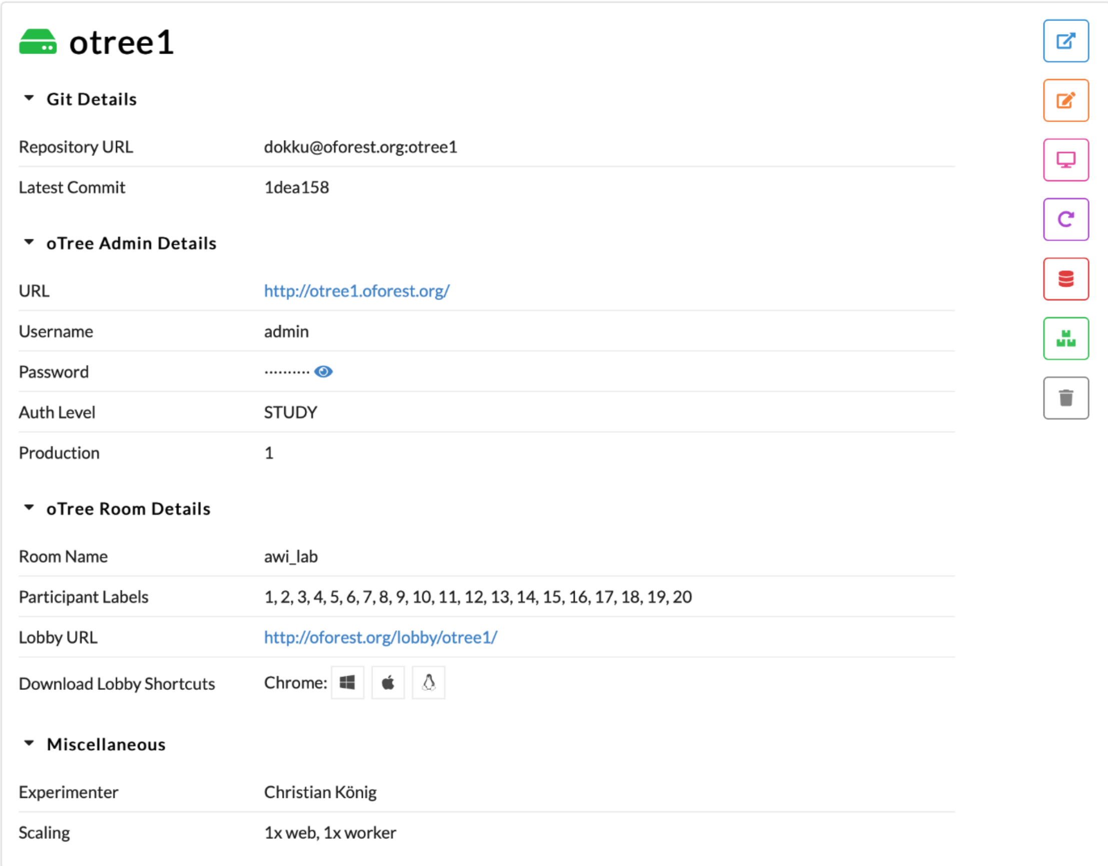
Notes: This shows the first screen that (super-)users see after logging in. Quick navigation icons are located in the top right (wrench is super-user only). Instances, which the current user has access to, are shown in a grid below. Each instance is represented by a box, showing a status icon (green = ready to run experiments: orange = ready to upload experiments), its name as well as the assigned experimenter's name. Super-users also have the option to set up a new oTree instance.

oTree Manager comes with a set of management features for experimenters: With the click of a button, they can (re-)set the password of oTree's web interface, restart the webserver, and reset the database (Figure 7.2). All of these actions would traditionally require command line interaction. Furthermore, they can integrate oTree's rooms feature with oTree Manager. If they do, oTree Manager automatically sets up a *Lobby* which can be opened on all client computers before the experimental session begins. It allows participants to signal that they are ready for the experiment to begin by clicking a button after having taken their seats (Figure 7.3). This step ensures that only those client computers show up on oTree's room management page, which actually have participants sitting in front of them. With this feature, experimenters can simply start all clients in the laboratory and direct the browsers to the Lobby's URL, irrespective of how many participants actually show up for the session. They do not need to worry about closing browser windows on unused client computers before starting the session.[87] oTree Manager provides desktop shortcuts for download, which start the Google Chrome (or Chromium) browser in kiosk mode and direct it to the lobby page for each participant label set up in oTree.

[87] There might be fewer participants present then client computers turned on due to participants not showing up to the session unexcused, for example. If the unused computers were pointed directly at oTree's room waiting page for session creation, the browsers would have to be closed manually in order not to have an incorrect number of clients connected when starting the actual experimental session.

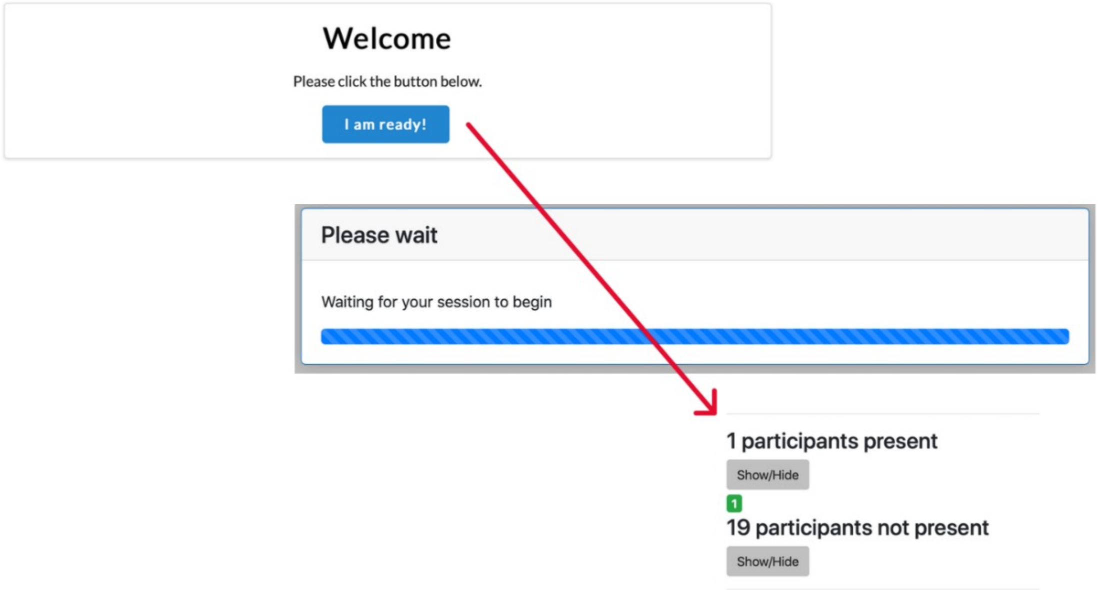
At any given time, experimenters can also monitor the state of their oTree instance. The web interface clearly communicates instance status (green = ready to run experiments / orange = ready to upload experiments), Git repository URL and credentials, oTree server URL, admin password, and current room setup, as well as the number of web and (timeout-)worker processes ready to handle incoming requests.

Figure 7.2: Detail View of an oTree Instance



Notes: This example detail view shows information for instance “otree1”. The green icon indicates that it is completely set up and is ready for experiments. The four foldable sections show various details regarding the Git repository, the oTree web interface, the Lobby configuration, and additional experimenter information. On the right, there are buttons for going to the oTree web interface, setting the admin password, restarting the instance, resetting the database, adjusting the scaling (super-users only) and deleting the installation (super-users only).

Figure 7.3: Lobby Feature Schematic



Notes: The schematic shows sections of three screens. The top left screen is the welcoming screen, which is shown to participants when they first sit down at their client computer. Once they confirm their readiness by clicking the blue button, they are forwarded to oTree's room waiting page (screen in the middle). Only now does the participant appear in oTree's room administration interfaces as being present (bottom right screen).

7.2.2 Laboratory Managers

Generally, laboratory managers will use so-called super-user accounts on oTree Manager's web interface. Super-user accounts behave like regular experimenter accounts, but come with more rights (and responsibilities). Super-user accounts allow the creation and management of individual user accounts for experimenters as well as oTree server instances. Super-users can create user accounts by providing a name for identification, a user name to login, and an e-mail address for communication. User accounts can also be promoted to super-user accounts such that the respective users have access to all features of oTree Manager. This is especially handy if more than one person manages the laboratory. Of course, user accounts can also be deleted.

Super-users have access and management rights to all oTree instances, irrespective of which user the instance belongs to. In terms of functionality, they can generally do everything a regular user can do and more. Specifically, super-users may change the numbers of web and (timeout-)worker processes that are started for each instance. That is, they can scale-up individual instances to be able to handle a larger number of simultaneous requests.[88] They can also delete oTree instances.

In its default configuration, oTree Manager automatically sets up sub-domains for the oTree instances. This keeps oTree server URLs short and easy to remember. It also allows using a single SSL certificate (with a wildcard entry for sub-domains) for all oTree instances running on the server. This keeps the effort required to add transport layer encryption to oTree instances to a minimum and enables rapid deployment of oTree instances suitable for integration with Amazon Mechanical Turk.

7.3 Architecture

7.3.1 Overview

Behind the scenes, oTree Manager creates a new oTree installation for each instance. Each instance comes with its own production-ready PostgreSQL and Redis database servers and provides Git access for easy deployment. By design, instances are completely independent of each other and thus changes to one instance cannot affect other instances.

On the lowest level, oTree Manager uses Docker to containerize and isolate oTree webservers and databases. Docker is a tool that allows software to be wrapped into standardized units (containers) which include most of the dependencies they need to run. These can easily be distributed, automatically set up, and cleanly separate their contents from each other and the host system. Importantly, containers are much more resource efficient than virtual machines. Each virtual machine comes with its own operating system, kernel, and software libraries. Often, it also has a pre-defined, fixed allocation of hardware resources. Containers, in contrast, run on top of the host's operating system. They can share its kernel, libraries, and hardware resources while maintaining a high level of separation. This reduces overhead and results in quicker start-up times and a higher number of 'guests' that can be run on the same host computer.

oTree Manager relies on Dokku, which is a lightweight, open source Platform as a Service (PaaS) implementation. It serves as the fabric between the user-facing web interface and the lower level Docker containers. Dokku manages containers, handles access rights and provides the Git repository functionality. In addition, it provides Heroku

[88] This is ultimately limited by the performance of the server oTree Manager is running on as well as the number of instances in use.

compatibility, which simplifies deployment for experimenters, as they do not need to make any changes to their local oTree development installation before pushing their code into production. At last, Dokku also handles sub-domain creation for the individual instances.

The web interface is built using the Django web framework, which is also the basis for oTree itself. Django is extended by Django Channels to run background processes (i.e. interface with Dokku) and provide near real-time feedback to users via Websocket notifications. Semantic-UI provides consistent and intuitive user interface elements.

The web interface is built using the Django web framework, which is also the basis for oTree itself. Django is extended by Django Channels to run background processes (i.e. interface with Dokku) and provide near real-time feedback to users via Websocket notifications. Semantic-UI provides consistent and intuitive user interface elements.

7.3.2 Details

The details of the processes taking place behind the scenes are described best by walking through a standard operation. The operation we take a closer look at is the creation of a new oTree instance with subsequent deployment of an experiment. The process is initiated by a super-user, who provides a name for the instances and assigns it to an existing experimenter account using the web interface. oTree Manager makes sure that instances names are both unique and URL safe.

If these conditions are met, multiple tasks are sent to Dokku: 1) A new empty app container is created which comes with a Git repository interface and is ready for Heroku-style deployments. 2) Two containers for PostgreSQL and Redis databases are created and linked to the app. 3) The assigned experimenter is granted proper access rights. Each of these steps is run as a separate task in the background using Django Channels. Once these tasks succeed or fail, they trigger notifications, which are shown on the logged-in super-user's interface. To facilitate this, a Websocket connection between the super-user's browser window and the oTree Manager server is kept open. Websockets are an efficient way to allow near real-time communication between server and client without the need for the browser to reload the page or send repeated queries to the server in the background.

Once the background tasks have completed, the new oTree instance shows up in the dashboard as *ready for deployment* (orange icon). Its detail view prominently shows the

Git URL which is used to transfer the experiments to the instance. The assigned experimenter can add this URL as a new remote repository to their existing Git repository used during development. Once the remote is set up, the experimenter can simply push the oTree project to the server instance. As soon as such a push event is picked up by the Dokku app, the oTree container is first re-built from scratch including its dependencies and then deployed.

At this point, if the instance has been deployed successfully, the state of the oTree container switches from *ready for deployment* (orange icon) to *ready for use* (green icon). For easy verification, the detail view now shows the currently deployed Git commit identifier. It also changes the presentation to focus on details of the deployed oTree instance such as its URL and the currently set oTree web interface admin password. Note that it is possible to change the password or adjust the number of processes even if the experiment has not been deployed, yet. Thus, super-users can pre-configure instances for experimenters, if they desire.

Actions such as restarting or deleting the instances, resetting the database, changing the admin password or scaling the number of worker processes work in a very similar way. User interactions trigger Dokku tasks, which are run in the background through Django Channels. These in turn notify the user once they have completed.

A complete manual, which includes user guides for both experimenters and super-users, as well as detailed installation instructions, is available on Read the Docs.[89] The source code of oTree Manager is published under the MIT License and available on GitHub.[90] In the spirit of open source, everyone is invited to contribute to the continued development of oTree Manager and its documentation.

7.4 Conclusion

oTree Manager makes it easy to set up, run, and manage multiple, production-ready oTree instances on a single machine. It comes with an intuitive web interface, which makes oTree installation management easier for both experimenters and laboratory managers. Instances managed through oTree Manager are completely independent from each other, come pre-configured for production use, and provide a handy Lobby feature

[89] http://docs.otree-manager.com
[90] https://github.com/otree-manager/otree_manager

for use in experimental laboratories. Using individual Docker containers for separate instances reduces resource requirements compared to traditional virtual machine setups and speeds-up initial deployment.

Glossary

Cloud-hosted	Typically, web services run on dedicated server hardware in data centers. This is expensive and does not easily scale to growing resource demands. Cloud-hosting runs web services in a virtualized environment which behaves like a single server, but can actually use resources from multiple machines and thus easily scale to growing performance demands. Multiple virtualized servers commonly share hardware resources which increases cost efficiency.
Dependencies	A key principle in software development is "don't repeat yourself" (DRY). Consequently, many software packages rely on other, supporting software to provide their functionality. These other required packages are commonly called dependencies.
Deployment	During its development process, software passes through various stages. These include actual development, testing, and finally deployment. Deploying a software means putting it into real-world use. For software to be considered ready for deployment it typically has to pass testing and quality control and be (mostly) free of known issues.
SSH (keys)	SSH stands for Secure Shell which is a network protocol for encrypted communication over unencrypted connections. It is typically used to log in to remote computers through a command line interface. SSH supports username and password credentials as well as identification and authentication through public-key cryptography. The latter is more secure and more convenient as it does not require the users to create and remember secure passwords.
Django	Django is an open-source web framework written in Python, which makes it easy to develop dynamic, data-driven websites. https://www.djangoproject.com
Django-Channels	Django-Channels is an extension of Django, which enables the use of more communication protocols. These are required, for example, for near real-time communication between browsers and web servers. https://channels.readthedocs.io
Docker	Docker bundles software into containers. These containers can be easily distributed and run on top of many common operating systems. Containers bring their own dependencies, tools, and libraries, but share the kernel with the host system. This allows them to be more resource efficient than other virtualization techniques. https://www.docker.com

Dokku	Dokku is a small Platform as a Service implementation. As such, it calls itself “mini-Heroku”, because it provides a user experience similar to Heroku. At its core, it is a collection of software scripts, which tie together Git repositories and Docker containers. http://dokku.viewdocs.io/dokku/
Git (repository / remote)	Git is a free and open-source version control system. A repository contains all files of a project, which are under version control. Git allows mirroring project repositories across multiple machines. A mirror on a different computer is typically called a “remote”. https://git-scm.com
GitHub	GitHub is an online platform, which hosts Git repositories. It provides a web interface to the repositories and augments them with useful collaboration features such as issue trackers, wikis, and team management. https://www.github.com
Heroku	Heroku is a cloud Platform as a Service provider. It makes it easy and quick to host websites and web services written in a multitude of languages. It is one of the recommended ways to deploy oTree experiments to production if an experimenter or lab is not running their own dedicated server infrastructure. https://www.heroku.com
Kernel	The Kernel is a central component of each operating system. Almost all input and output requests from software pass through it on their way to the different hardware components.
MIT License	The MIT License allows software to be used for private and commercial purposes, allows it to be distributed, and modified. It limits liability and offers no warranty. A copyright notice as well as the permission statement need to be included if substantial portions of the licensed software are copied or re-used. https://opensource.org/licenses/MIT
Multi-user support	Software can be designed for use by a single user or for use by multiple users. The latter requires some form of user- and credentials management to be included. It is typically expected that software, which supports multiple users, keeps their work separate to prevent interference and promote data privacy and protection.
Platform as a Service	A Platform as a Service (PaaS) is software, which reduces the complexity of developing and running web services. These platforms are typically pre-configured for common deployment patterns and provide easy access to databases and storage providers. Often they

	run on cloud infrastructure, which enables them to scale quickly and transparently to growing resource demands.
Port	Ports are endpoints of computer network communications. Ports can be imagined as doors to the computer. As such, ports can be closed and locked or they can be open. Software which communicates over the network “listens” to assigned ports for incoming connections or sends own messages addressed to ports on the receiving end. Ports are associated with an IP address and the network protocol to be used. Typically, they are appended to the remote address with a colon (e.g., :8000).
PostgreSQL	PostgreSQL is an object-relational database software. It stores data in forms of tables in which columns correspond to variable names and rows to individual entries. The Structured Query Language (SQL) is used to manage data contained in the database, giving it its name. https://www.postgresql.com
Read the Docs	Read the Docs is an open-source software documentation hosting platform. http://readthedocs.org
Redis	Redis is an in-memory key-value database. Data storage is tailored towards fast retrieval. As such, there is no pre-defined structure of data as in typical object-relational databases. Each record can contain its own individual collection of fields. https://redis.io/
SSL / TSL	Secure Sockets Layer (SSL) is the predecessor to Transport Layer Security (TLS). Both are network protocols for encrypted communication and are used to provide privacy and data integrity between clients and servers.
URL	URL stands for Uniform Resource Locator. URLs are more commonly known as web addresses.
Virtual environment	On computer systems used by multiple users one goal is to keep users’ data separate from each other and private. While users may share some resources such as installed software, it can be beneficial to also have software separation. This way, each user can keep their own configuration and set of dependencies as well as specific versions, which might otherwise lead to conflicts between users in shared environments. Virtual environments make this separation possible.
Virtual Machine	A virtual machine is a virtualized (simulated) computer running on top of a host’s operating system. It typically emulates a complete set of hardware, requiring the installation of its own, separate operating system to be useful. While there is a large degree of separation between different virtual machines running on the same

	host, they also come at a large resource overhead, because each one brings their own, complete operating system.
Web/worker processes	oTree runs multiple types of processes. Web processes handle incoming network requests and serve pages. Worker processes make sure that timeouts occur, even if a participant in the experiment has closed their browser. Multiple web processes can improve performance if many requests are to be handled simultaneously, for example in large online (Amazon Mechanical Turk) experiments.
Websocket	Websocket is a network communication protocol. It allows sending and receiving messages on the same ports typically in use by web servers. Because the connection between server and client is kept open after initialization, it enables near real-time communication between servers and clients. It can be used to implement chats for example.

Chapter 8

Conclusion

What can we take away from this accumulation of research projects? I see two big themes that emerge from the individual articles. First, there is the aspect of how communication affects decision making in financial settings. Chapter 2 demonstrates that even if communication is highly structured and uses well-established terminology, perception about its meaning can differ starkly between individuals. As we show, these differences have real consequences for decision makers and those affected. Agents might try to follow their principals' wishes, but still unintendedly fail in reaching a suitable outcome. Yet, the same article also highlights that communication, even if it is as noisy as the perception of investment profile terminology, still has the power to guide decisions. This observation is mirrored by our findings in chapter 5. Here, we see that deliberately changing the precision of information available to depositors affects their decisions and can have far-reaching consequences for the interconnected system of financial institutions.

At the same time, both studies also tell us that contextual factors are a second big determinant of how information affects behavior. In chapter 2, we find that advisors serving multiple clients simultaneously differentiate more strongly between clients with different tastes than advisors who only observe the preferences of a single customer. In the bank run study, we see that information that is more precise has differential effects for banks with strong and weak fundamentals. We also observe that bank stability information is adequately transferred to other institutions if they are interconnected. Yet, disclosed information is not applied in decisions regarding other institutions, if interbank linkages are non-existent. Clearly, context is key.

However, it is not just decision makers who are affected by information availability and precision. The three experiments of chapter 3 speak a clear language: Evaluators of financial decisions often take outcome information into account, even if it is provably unrelated to decision quality. While the observation of outcome bias is not unexpected, its strength and robustness to monetary incentives in financial settings is still surprising. Notably, the effects of outcome information on decision evaluations also depend on contextual factors: We observe that positive outcomes exert a much stronger pull on evaluations than negative ones. It appears that the evaluator's mindset affects how outcome information is processed.

Taken together, these findings impressively demonstrate how contextual factors, information precision, and information perception work together in shaping behavior of decision makers in financial settings.

The second more general topic that my thesis speaks to is the issue of generalizability of research findings and their robustness to different influences. Chapters 2, 4, and 6 can be read in this context. In chapter 2, we observe that the finding of Foerster et al. (2017) that financial advisors' own preferences also affect the composition of their clients' portfolios appears to be qualitatively robust to substancial changes in research methodology, sample selection, and even the cultural and institutional background. While the strengths of the different determinants of behavior might differ between laboratory and real-life situations, there is no doubt that the fundamental mechanisms are captured in both cases.

Yet, there are other cases in which initial observations might be more context dependent. As reported in chapter 6, despite only changing the participant sample from financial professionals to students while keeping all other aspects identical, we are unable to find evidence of countercyclical risk aversion, while Cohn et al. (2015) report large effects. It remains inherently unclear, which factors determine how well results generalize and how robust findings are to perturbations in the decision making context. Once again, it has become evident that tests of generalizability and outright replications are a necessary and valuable contribution to research in economics.

Connected to generalizability and the question of how much we actually learn from single economic experiments is the issue of modeling the decision environment. While many real-life decisions involving uncertainty about outcomes are modeled as decisions under risk, it is unclear whether this is actually fitting or just a convenient choice. As we show in chapter 4, modeling and designing a decision situation around ambiguity rather than risk can bring about very different behavioral responses by participants. Recall that we do not find any differences in the ambiguity attitudes expressed in participants' decisions for themselves, for others, and for others under accountability. This is in stark contrast to the results reported for very similar decisions with uncertainty modeled as risk in Pollmann et al (2014).

Clearly, it is advisable to study the same research questions using different methods, modeling techniques, and under different sets of simplifying assumptions to make sure the results are actually robust and sufficiently generalizable. Otherwise, we might end up in the unfortunate situation of formulating policy advice based on an insufficient understanding of the issue at hand.

References

Ahnert, T. and Georg, C-P. (2018). Information contagion and systemic risk. *Journal of Financial Stability*, *35*, 159–171.

Alevy, J.E., Haigh, M.S., and List, J.A. (2007). Information Cascades: Evidence from a Field Experiment with Financial Market Professionals. *Journal of Finance*, *62*(1), 151–180.

Andersson, O., Holm, H.J., Tyran, J.R., and Wengström, E. (2014). Deciding for others reduces loss aversion. *Management Science*, *62*(1), 29–36.

Arifovic, J. and Jiang, J.H. (2014). Do Sunspots Matter? Evidence from an Experimental Study of Bank Runs. Bank of Canada Working Paper.

Arifovic, J., Jiang, J.H., and Xu, Y. (2013). Experimental evidence of bank runs as pure coordination failures. *Journal of Economic Dynamics and Control*, *37*(12), 2446–2465.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2018) Incentives in Experiments: A Theoretical Analysis. *Journal of Political Economy, 126*(4), 1472-1503.

Baron, J. and Hershey, J.C. (1988). Outcome bias in decision evaluation. *Journal of Personality and Social Psychology*, *54*(4), 569−579.

Baumeister, R.F., Bratslavsky, E., Finkenauer, C., and Vohs, K.D. (2001). Bad is Stronger than Good. *Review of General Psychology*, *5*, 323–370.

Bertrand, M. and Mullainathan, S. (2001). Are CEOs rewarded for luck? The ones without principals are. *Quarterly Journal of Economics*, *116*(3), 901−932.

Blank, H., Diedenhofer, B., and Musch, J. (2015). Looking back on the London Olympics: Independent outcome and hindsight effects in decision evaluation. *British Journal of Social Psychology*, *54*, 798–807.

Bock, O., Baetge, I., and Nicklisch, A. (2014). hroot – Hamburg registration and organization online tool. *European Economic Review*, *71,* 117–120.

Bolton, G.E. and Ockenfels, A. (2010). Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment. *American Economic Review*, *100*(1), 628–633.

Bosman, R. and van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *Economic Journal, 112*(476), 147–169.

Bouvard, M., Chaigneau, P., and Motta, A.D. (2015). Transparency in the Financial System: Rollover Risk and Crises. *The Journal of Finance*, *70*(4), 1805–1837.

Bradbury, M., Hens, T., and Zeisberger, S. (2015). Improving Investment Decisions with simulated experience. *Review of Finance*, *19*, 1019–1052.

Bradley, M.M., and Lang, P.J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59.

Brown, M., Trautmann, S.T., and Vlahu, R.E. (2017). Understanding Bank-Run Contagion. *Management Science*, *63*, 2272–2282.

Cachon, G.P. and Camerer, C.F. (1996). Loss-avoidance and forward induction in experimental coordination games. *Quarterly Journal of Economics*, *111*, 165–194.

Callen, M., Isaqzadeh, M., Long, J.D., and Sprenger, C. (2014). Violence and Risk Preference: Experimental Evidence from Afghanistan. *American Economic Review*, *104*(1), 123–148.

Calomiris, W.C. and Jaremski, M. (2016). Deposit insurance: Theories and facts. NBER Working Paper No. 22223.

Carlsson, H. and Van Damme, E. (1993). Global games and equilibrium selection. *Econometrica*, *61*(5), 989–1018.

Casal, S., Ploner, M., and Sproten, A. (2019). Fostering the best execution regime: An experiment about pecuniary sanctions and accountability in fiduciary money management. *Economic Inquiry*, *57*(1), 600–616.

Chakravarty, S., Fonseca, M.A., and Kaplan, T.R. (2014). An Experiment on the Causes of Bank Run Contagions. *European Economic Review*, *72*, 39–51.

Chakravarty, S., Harrison, G.W., Haruvy, E.E., and Rutström, E.E. (2011). Are you risk averse over other people's money? *Southern Economic Journal*, *77*(4), 901–913.

Charness, G. and Jackson, M.O. (2009). The role of responsibility in strategic risk-taking. *Journal of Economic Behavior & Organization*, *69*(3), 241–247.

Charness, G., Karni, E., and Levin, D. (2010). On the conjunction fallacy in probability judgement: New experimental evidence regarding Linda. *Games and Economic Behavior*, *68*(2), 551–556.

Charness, G., Karni, E., and Levin, D. (2013). Ambiguity attitudes and social interactions: An experimental investigation. *Journal of Risk and Uncertainty*, *46*(1), 1–25.

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

Chen, Y. (1999). Banking panics: The role of first-come, first-served rule and information externalities. *Journal of Political Economy*, *107*(5), 946–968.

Cipriani, M., Guarino, A., Guazzarotti, G., Tagliati, F., and Fischer, S. (2018). Information Contagion in the Laboratory. *Review of Finance*, *22*(3), 877–904.

Cohn, A., Engelmann, J., Fehr, E., and Maréchal, M.A. (2015). Evidence for Countercyclical Risk Aversion: An Experiment with Financial Professionals. *American Economic Review*, *105*(2), 860–885.

Cordes, H. and Dierkes, M. (2017). About depression babies and red diaper babies: Do macroeconomic experiences affect everybody's risk taking in the same way? *Journal of Behavioral and Experimental Finance*, *13*, 24–27.

Cordes, H., Nolte, S., and Schneider, J.C. (2017). On the dynamics and drivers of countercyclical risk aversion. SSRN Working Paper no. 2918953.

Crusius, J., van Horen, F., and Mussweiler, T. (2012). Why process matters: A social cognition perspective on economic behavior. *Journal of Economic Psychology*, *33*, 677–685.

Cushman, F., Dreber, A., Wang, Y., and Costa, J. (2009). Accidental Outcomes Guide Punishment in a "Trembling Hand" Game. *PLoS ONE, 4*(8), e669.

Davis, D.D. and Reilly, R.J. (2016). On Freezing Depositor Funds at Financially Distressed Banks: An Experimental Analysis. *Journal of Money, Credit and Banking*, *48*(5), 898–1017.

De Oliveira, A.C.M., Smith, A., and Spraggon, J. (2017). Reward the lucky? An experimental investigation of the impact of agency and luck on bonuses. *Journal of Economic Psychology*, *62*, 87–97.

Demirguc-Kunt, A., Kane, E., and Laeven, L. (2015). Deposit insurance around the world: A comprehensive analysis and database. *Journal of Financial Stability*, *20*, 155–183.

Demirguc-Kunt, A., Karacaovali, B., and Laeven, L. (2005). Deposit insurance around the world: A comprehensive database. Policy Research Working Paper No. 3628, World Bank.

Diacon, S. (2004). Investment risk perceptions: Do consumers and advisers agree? *International Journal of Bank Marketing*, *22*(3), 180–199.

Diamond, D.W. and Dybvig, P.H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, *91*(3), 401–419.

Dijk, O. (2017). Bank run psychology. *Journal of Economic Behavior & Organization, 144*, 87–96.

Egan, M., Matvos, G., and Seru, A. (2019). The Market for Financial Adviser Misconduct. *Journal of Political Economy*, *127*(1), 233–295.

Emerson, G. B., Warme, W.J., Wolf, F.M., Heckman, J.D., Brand, R.A., and Leopold, S.S. (2010). Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Archives of Internal Medicine*, *170*(21), 1934–1939.

Eriksen, K.W. and Kvaløy, O. (2010). Do financial advisors exhibit myopic loss aversion? *Financial Markets and Portfolio Management*, *24*(2), 159–170.

European Central Bank. (2015). Greece, Outstanding amounts at the end of the period (stocks), MFIs excluding ESCB reporting sector – Deposit liabilities, Total maturity, All currencies combined - Euro area (changing composition) counterpart, Non-MFIs excluding central government sector, denominated in Euro, data Neither seasonally nor working day adjusted (No. BSI.M.GR.N.A.L20.A.1.U2.2300.Z01.E).

Fehr, E. and Schmidt, K.M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, *114*(3), 817–868.

Feng, L. and Seasholes, M. (2005). Do Investor Sophistication and Trading Experience Eliminate Behavioral Biases in Financial Markets? *Review of Finance*, *9*(3), 305–351.

Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Flannery, M.J. (1998). Using market information in prudential bank supervision: A review of the US empirical evidence. *Journal of Money, Credit and Banking*, *30*, 273–305.

Foerster, S., Linnainmaa, J.T., Melzer, B.T., and Previtero, A. (2017). Retail Financial Advice: Does One Size Fit All? *Journal of Finance*, *72*(4), 1441–1482.

Füllbrunn, S. and Luhan, W.J. (2017). Am I My Peer's Keeper? Social Responsibility in Financial Decision Making. University of Portsmouth Working Papers in Economics & Finance.

Gächter, S., Johnson, E.J., and Herrmann, A. (2010). Individual-level loss aversion in riskless and risky choices. CeDEx discussion paper series, No. 2010-20.

Garratt, R. and Keister, T. (2009). Bank runs as coordination failures: An experimental study. *Journal of Economic Behavior & Organization*, *71*(2), 300–317.

Gilad, D. and Kliger, D. (2008). Priming the Risk Attitudes of Professionals in Financial Decision Making. *Review of Finance*, *12*, 567–586.

Glaser, M., Iliewa, Z., and Weber, M. (2019). Thinking about prices versus thinking about returns in financial markets. *Journal of Finance*, *forthcoming*.

Glasserman, P. and Young, P. (2016). Contagion in Financial Networks. *Journal of Economic Literature*, *54*(3), 779–831.

Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, *112*(2), 631–645.

Goldstein, I. and Leitner, Y. (2018). Stress tests and information disclosure. *Journal of Economic Theory*, *177*, 34–69.

Goldstein, I. and Yang, L. (2017). Information Disclosure in Financial Markets. *Annual Review of Financial Economics*, *9*,101–125.

Goldstein, I. and Sapra, H. (2014). Should banks' stress test results be disclosed? An analysis of the costs and benefits. *Foundations and Trends in Finance*, *8*(1), 1–54.

Grable, J. and Lytton, R.H. (1999). Financial risk tolerance revisited: the development of a risk assessment instrument. *Financial Services Review*, *8*(3), 163–181.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125.

Gurdal, M.Y., Miller, J.B., and Rustichini, A. (2013). Why Blame? *Journal of Political Economy*, *121*(6), 1205−1247.

Haigh, M.S. and List, J.A. (2005). Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis. *Journal of Finance*, *60*(1), 523–534.

Hallahan, T.A., Faff, R.W., and McKenzie, M.D. (2004). An empirical investigation of personal financial risk tolerance. *Financial Services Review*, *13*(1), 57–78.

Hegglin, R. (2015). System Stability and (Bad) Experience: An Experimental Study of Banking Crises. SSRN Scholarly Paper No. ID 2620157.

Hershey, J.C. and Baron, J. (1992). Judgment by outcomes: When is it justified? *Organizational Behavior and Human Decision Processes*, *53*(1), 89–93.

Hsee, C.K. and Weber, E.U. (1997). A fundamental prediction error: Self–others discrepancies in risk preference. *Journal of Experimental Psychology: General*, *126*(1), 45–53.

Ifcher, J. and Zarghamee, H. (2018). Behavioral Economic Phenomena in Decision-Making for Others. IZA Discussion Paper No. 11946.

Irtel, H. (2008). The PXLab Self-Assessment-Manikin Scales available from http://irtel.uni-mannheim.de/pxlab/demos/index_SAM.html. Part of the software package PXLab: The Psychological Experiments Laboratory, Version 2.1.11. University of Mannheim. Available from http://www.pxlab.de.

Iyer, R. and Puri, M. (2012). Understanding bank runs. *American Economic Review*, *102*(4), 1414–1445.

Iyer, R. and Peydro, J.-L. (2011). Interbank contagion at work: Evidence from a natural experiment. *Review of Financial Studies*, *24*(4), 1337–1377.

Karas, A., Pyle, W., and Schoors, K. (2013). Deposit insurance, banking crises and market discipline: Evidence from a natural experiment on deposit flows and rates. *Journal of Money, Credit and Banking*, *45*(1), 179–200.

Kaufman, G.G. (1994). Bank contagion: A review of the theory and evidence. *Journal of Financial Services Research*, *8*(2), 123–150.

Kaufmann, C., Weber, M., and Haisley, E. (2013). The role of experience sampling and graphical displays on one's investment risk appetite. *Management Science*, *59*(2), 323–340.

Kaustia, M., Alho, E., and Puttonen, V. (2008). How Much Does Expertise Reduce Behavioral Biases? The Case of Anchoring Effects in Stock Return Estimates. *Financial Management, 37*(3), 391–411.

Kirchler, M., Lindner, F., and Weitzel, U. (2018). Rankings and risk-taking in the finance industry. *Journal of Finance*, *73*, 2271–2302.

Kiss, H.J., Rodriguez-Lara, I., and Rosa-García, A. (2016). Think twice before running! Bank runs and cognitive abilities. *Journal of Behavioral and Experimental Economics*, *64*, 12–19.

Kling, L., König-Kersting, C., and Trautmann S.T. (2019). Investing for Others: Principals‘ vs. Agents' Preferences. Mimeo (Chapter 2).

Klos, A., and Sträter, N. (2013). How Strongly Do Players React to Increased Risk Sharing in an Experimental Bank Run Game? QBER Discussion Paper No. 6/2013.

König-Kersting, C. and Trautmann, S.T. (2016). Ambiguity Attitudes in Decisions for Others. *Economics Letters*, *146*, *1*26–129.

König-Kersting, C. and Trautmann, S.T. (2018). Countercyclical Risk Aversion: Beyond Financial Professionals. *Journal of Behavioral and Experimental Finance*, *18*, 94–101.

Lefebvre, M., Vieider, F.M., and Villeval, M.C. (2011). The ratio bias phenomenon: fact or artifact? *Theory and Decision*, *71*(4), 615–641.

Lefgren, L., Platt, B., and Price, J. (2015). Sticking with What (Barely) Worked: A Test of Outcome Bias. *Management Science*, *61*(5), 1121–1136.

Leiner, D.J. (2014). Software: SoSci Survey (Version 2.5.00-i), available at https://soscisurvey.de.

Leitner, Y. (2014). Should Regulators Reveal Information About Banks? Federal Reserve Bank of Philadelphia Business Review, Third Quarter.

Lipshitz, R. (1989). "Either a medal or a corporal": The effects of success and failure on the evaluation of decision making and decision makers. *Organizational Behavior and Human Decision Processes*, *44*(3), 380–395.

List, J. (2003). Does market experience eliminate market anomalies? *The Quarterly Journal of Economics*, *118*(1), 41–71.

Loewenstein, G. (1996). Out of control: Visceral influences on behavior. *Organizational Behavior and Human Decision Processes*, *65*(3), 272–292.

Lusardi, A. and Mitchell O.S. (2011). Financial literacy around the world: an overview. *Journal of Pension Economics and Finance*, *10*(4), 497–508.

Malmendier, U. and Nagel, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking? *Quarterly Journal of Economics*, *126*(1), 373–416.

Marshall, G. W. and Mowen, J. C. (1993). An experimental investigation of the outcome bias in salesperson performance evaluations. *Journal of Personal Selling & Sales Management*, *13*(3), 31–47.

Martinez Peria, M.S. and Schmukler, S.L. (2001). Do depositors punish banks for bad behavior? Market discipline, deposit insurance, and banking crises. *Journal of Finance*, *56*(3), 1029–1052.

Montinari, N. and Rancan, M. (2018). Risk taking on behalf of others: the role of social distance. *Journal of Risk and Uncertainty*, *57*(1), 81–109.

Morris, S. and Shin, H.S. (2002). Social Value of Public Information. *American Economic Review*, *92*(5), 1521–1534.

Müller, D. and Weinschenk, P. (2015). Rater Bias and Incentive Provision. *Journal of Economics & Management Strategy*, *24*(4). 833–862.

Murata, A., Nakamura, T., Matsushita, Y., and Moriwaka, M. (2015). Outcome bias in decision making on punishment or reward. *Procedia Manufacturing*, *3*, 3911–3916.

Mutual Fund Dealers Association of Canada (2014). MFDA Discussion Paper on the Use of Investor Questionnaires available from: http://mfda.ca/bulletin/Bulletin0611-C/

Nier, E.W. (2005). Bank stability and transparency. *Journal of Financial Stability*, *1*, 342–354.

Nier, E.W. and Baumann, U. (2006). Market discipline, disclosure and moral hazard in banking. *Journal of Financial Intermediation*, *15*, 332–361.

Oechssler, J. and Roomets, A. (2014) Unintended Hedging in Ambiguity Experiments. *Economics Letters, 122,* 243-246.

Pahlke, J., Strasser, S., and Vieider, F.M. (2015). Responsibility effects in decision making under risk. *Journal of Risk and Uncertainty*, *51*(2), 125–146.

Pollmann, M.M.H., Potters, J., and Trautmann, S.T. (2014). Risk taking by agents: The role of ex-ante and ex-post accountability. *Economics Letters*, *123*(3), 387–390.

Polman, E. (2012). Self–other decision making and loss aversion. *Organizational Behavior and Human Decision Processes*, *119*(2), 141–150.

Popper, K.R. (1974). *Conjectures and Refutations*. 5^{th} Ed. London, Routledge and Kegan Paul.

Prean, N. and Stix, H. (2011). The effect of raising deposit insurance coverage in times of financial crisis – Evidence from Croatian microdata. *Economic Systems*, *35*(4), 496–511.

Ratner, R.K. and Herbst, K.C. (2005). When good decisions have bad outcomes: The impact of affect on switching behavior. *Organizational Behavior and Human Decision Processes*, *96*(1), 23–37.

Reynolds, D.B., Joseph, J., and Sherwood, R. (2009). Risky shift versus cautious shift: determining differences in risk taking between private and public management decision-making. *Journal of Business and Economics Research*, 7(1), 63–78.

Roszkowski, M.J. and Grable, J.E. (2005). Estimating risk tolerance: The degree of accuracy and the paramorphic representations of the estimate. *Journal of Financial Counseling and Planning*, *16*(2), 29–47.

Rud, O.A., Rabanal, J.P., and Horowitz, J. (2019). Does Competition Aggravate Moral Hazard? A Multi-Principal-Agent Experiment. Working Paper.

Rydval, O. and Ortmann, A. (2005). Loss avoidance as selection principle: Evidence from simple stag-hunt games. *Economics Letters*, *88*, 101–107.

Scheier M.F. and Carver, C.S. (1985). Optimism, coping, and Health: Assessment and Implications of Generalized Outcome Expectancies. *Health Psychology*, *4*(3), 219–247.

Scheier M.F., Carver. C.S., and Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology*, *67*(6), 1063–1078.

Schotter, A. and Yorulmazer, T. (2009). On the dynamics and severity of bank runs: An experimental study. *Journal of Financial Intermediation*, *18*(2), 217–241.

Sezer, O., Zhang, T., Gino, F., and Bazerman, M.H. (2016). Overcoming the outcome bias: Making intentions matter. *Organizational Behavior and Human Decision Processes*, *137*, 13–26.

Shakina, E. and Angerer, M. (2018). Coordination and communication during bank runs. *Journal of Behavioral and Experimental Finance*, *20*, 115–130.

Shapira, Z. and Venezia, I. (2001). Patterns of behavior of professionally managed and independent investors. *Journal of Banking & Finance*, *25*(8), 1573–1587.

Sirri, E.R. and Tufano, P. (1998). Costly Search and Mutual-Fund Flows. *Journal of Finance*, *53*(5), 1589–1622.

Slovic, P. (1987). Perception of Risk. *Science, 236*(4799), 280–285.

Trautmann, S.T. (2009). A tractable model of process fairness under risk. *Journal of Economic Psychology*, *30*(5), 803–813.

Trautmann, S.T. and Vlahu, R. (2013). Strategic loan defaults and coordination: An experimental analysis. *Journal of Banking & Finance*, *37*(3), 747–760.

Trautmann, S.T. and van de Kuilen, G. (2015). Ambiguity Attitudes, in: Keren, G., Wu, G. (Eds.), The Wiley Blackwell Handbook of Judgment and Decision Making. Blackwell, Chapter 3, 89−116.

Trautmann, S.T., Vieider, F.M., and Wakker, P.P. (2008). Causes of Ambiguity Aversion: Known versus Unknown Preferences. *Journal of Risk and Uncertainty*, *36,* 225−243.

Trevino, I. (2019). Informational channels of financial contagion. Working Paper.

Van Rooij, M., Lusardi, A., and Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, *101*(2), 449–472.

Weitzel, U., Huber, C., Huber, J., Kirchler, M., Lindner, F., and Rose, J. (2018). Bubbles and Financial Professionals. Innsbruck Working Papers in Economics and Statistics 2018-04.