Behavioral Foundations of Search, Matching, Teamwork, and Project Evaluation

Preferences and Constraints in Decision-Making

Timm Christian Opitz

Dissertation

Munich 2024

Max Planck Institute for Innovation and Competition
LMU Munich, Department of Economics

Behavioral Foundations of Search, Matching, Teamwork, and Project Evaluation

Preferences and Constraints in Decision-Making

Inaugural-Dissertation
zur Erlangung des Grades

Doctor oeconomiae publicae
(Dr. oec. publ.)

an der
Ludwig–Maximilians–Universität München

2024

vorgelegt von

Timm Christian Opitz

Referent: Prof. Dietmar Harhoff, Ph.D.
Korreferent: Prof. Dr. Florian Englmaier
Promotionsabschlussberatung: 31. Januar 2024

Datum der mündlichen Prüfung:	22. Januar 2024
Namen der Berichterstatter:	Prof. Dietmar Harhoff, Ph.D.
	Prof. Dr. Florian Englmaier
	Prof. Vojtěch Bartoš, Ph.D.

Acknowledgements

I am grateful that this dissertation reflects who I have become as a researcher. This would have been impossible without the support of Dietmar Harhoff. He encouraged me to try out own ideas, to pursue unconventional research projects, and to become an independent researcher early on. In combination with the provision of a safety net, the guidance when it came to the big decisions, and the funding of MPI allowed me to do exactly the kind of work that I am excited about. I am also grateful for the support I received from Florian Englmaier, both when it came to making a stay at the University of Toronto possible, to discussing research ideas, and to the provision of a supportive and collaborative environment within the group at LMU. I would like to thank Vojtěch Bartoš for completing my committee and being a true mentor in my academic journey.

My career path would most likely have been different without the influence of some people in the profession. Leandro Prados de la Escosura sparked my interest in an academic career during my undergraduate stay at UC3M, Martin Watzinger brought it back when it was about to get lost. Simeon Schudy and Tobias Schuwerk then succeeded in showing me how to combine my interests in economics and psychology – which is reflected in this thesis.

In large parts, this dissertation is a result of collaborative efforts. I have learned a lot through working with my co-authors over the last years. Therefore I want to thank all of them for the joint efforts, this would have been impossible without them. A special thanks goes to Christoph Schwaiger, with whom I had some of the most rewarding times in research. It has always been a pleasure to think through problems together, and discuss research as well as non-research related matters. However, it is not only the co-authors who were involved in the collaborative efforts. It is also research assistants, project managers, and supporting staff in Uganda, who worked with us on the different research projects, and to whom I am especially grateful.

The environment at the Max Planck Institute for Innovation and Competition has been key to this thesis – and this does not primarily refer to the great facilities and the support for research. Although not all of my research has been at the core of our institute's mission, I could always count on the help and support of my current and former colleagues. Thank you! Nevertheless, I would like to highlight Cristina Rujan as a colleague I could always turn to, and who has become one of my best friends during our time at MPI. Beyond the MPI, the Micro group at the LMU has been my second academic home during these past years. Thank you for welcoming me, thank you for being generous with your time when discussing research, thank you for all the good times.

I would probably not have completed my dissertation without having people around me who do not care too much about my research (and this dissertation). Most of my closest friends have only a vague idea of what I have been doing these years. Thank you for providing these environments –both in Munich and in Frankfurt– that allowed for distractions and helped me keep up with other priorities. I would like to especially thank my parents for letting me follow my own path without reservation, for always being supportive, and for never asking when I would finally submit this thesis.

I am proud of the work and effort that went into this dissertation. At the same time, the knowledge that it does not define me as a person, and that there is so much more, fills me with joy and happiness. I am incredibly grateful to always have you, Kathrin, by my side.[1]

[1]To Ida, if you ever read the first pages of this dissertation: I love you.

Contents

Preface

Understanding how people make decisions is key for designing economic policies. Traditional economic models tend to ignore that human decisions do not always follow purely profit-maximizing strategies under perfect information. In the last decades, behavioral economics has shed light on social dimensions of human decision-making, on constraints and biases of individuals, as well as on decision processes in the absence of perfect knowledge. One of the tools to analyze such questions are experiments. In principle, experiments allow the researcher to design tightly controlled environments to causally identify drivers of decision processes. At the same time, behavioral economists face the challenge of not diluting the rigor of economic thinking with an "anything goes" mentality, where phenomena with low economic and social relevance, limited external validity, and little chance of replication are studied.

But how can behavioral economics meaningfully contribute to policy? One measure of success of a scientific discipline is "*the extent to which it becomes the source of practical advice, solidly grounded in well-tested theory* [...]", as Al Roth noted (Roth, 1991). For behavioral economics, this entails at least four key lessons. First, tackle policy-relevant questions. Second, focus on generalizable mechanisms. Third, use economic theory to guide the design of experiments and the interpretation of results. Fourth, use methods that allow to create a solid knowledge base for policy-making. So it is a misconception to think that behavioral economics is only about understanding the psychological principles that govern economic interactions. It is at least as much about how well we can apply those principles to practical problems.

This dissertation focuses on preferences and constraints underlying individual decision-making. I investigate what makes consumers search inefficiently, why people match with whom, when teams collaborate effectively, and what prevents female entrepreneurs from succeeding. The first essay examines the effects of regret and perceived urgency on optimal search in a well-powered laboratory experiment. This contributes to policy debates about the need to regulate pressure selling in dynamic market environments such as online marketplaces. The second essay demonstrates that individuals have strong preferences to being matched with those who also want to be matched with them. We study implications for efficient team formation, and investigate the underlying mechanisms. The third essay investigates how interpersonal preferences affect modern teamwork. It provides insights into how teams should be optimally composed, and whether self-selection of team members may be detrimental to performance. The fourth essay analyzes constraints to the growth of female businesses in a low-income setting due to gendered access to finance. We investigate whether and why female business ideas are evaluated differently, and how the formation of entrepreneurial teams can overcome biased assessments.

These research projects are motivated by policy-relevant observations. In the search context, we directly address competition authorities' concerns about consumers being "pressurised by practices implying that they must act quickly to avoid missing out" (UK Competition and Markets Authority). To do so, we investigate the role of pressure tactics in inefficient search behavior. In the context of matching and collaboration, we tackle the crucial question of how to improve the formation and performance of teams. We provide evidence for why centralized mechanisms designed to form teams efficiently may fail to achieve their objective. We highlight the importance of considering that participants often prefer to interact with those who also want to interact with them. Then, I extend these by analyzing under which conditions self-selected teams may perform better or worse when faced with complex problem solving. In the context of entrepreneurial finance, we are motivated by the apparent mismatch between the potential of entrepreneurship as a key to economic development and the fact that most firms in developing countries tend to stay small and rarely grow beyond subsistence size. This pattern of stagnation is particularly pronounced for female entrepreneurs. We examine whether and why access to finance prevents women from realizing the potential of their entrepreneurial ideas and transforming their start-ups into successful businesses.

All projects tackle questions that have been difficult to answer with traditional econometric strategies due to a lack of suitable natural variation. Combining the experimental approach with economic theory allows me to identify concrete mechanisms that lead to the policy-relevant observations. I use the experimental toolkit from standard laboratory experiments to framed field experiments. The experimental procedures are pre-registered and thus conducive to the replicability of science. Theory guides the design of the experimental environments that ultimately allow me to control information sets and selection. This enables me to carefully study the mechanisms through which preferences and constraints affect individual decisions that turn into policy-relevant outcomes. These preferences and constraints include fear of regret, time pressure, interpersonal preferences, or gender biases. The causal identification of these mechanisms allows for the design of tailored policies to address them. To summarize, I am interested in the emergence of (at least seemingly) inefficient outcomes and the role that preferences and constraints play in their emergence.

Chapter I starts from the observation that perceptions of urgency and regret are common to many sequential search processes.[2] Online marketplaces such as flight booking sites are an example of this. By highlighting how few seats are available, how many people are looking for similar flights, or that fares are only guaranteed for a limited time, these websites induce two things. They create a sense of urgency to make a decision now, and they play on future feelings of regret if one does not buy on the spot. But it is not only sellers who pressure buyers with time limited offers and emphasize potential regret over missed opportunities; urgency and regret are present in many other settings as well. For example, job seekers face deadlines and anticipate (or experience) regret when rejecting or accepting offers, and investors who experience rapid price changes may regret missed opportunities to sell.

We theorize that regret and time pressure may explain the inefficiently short search lengths documented in the prior literature (e.g., Hey, 1987; Cox & Oaxaca, 1989; Einav, 2005). In a standard search that ends with a purchase, there is only one type of regret: regret for having searched too long. In the flight search example, a consumer would regret if they found a good deal early on, but decided to continue searching without finding a better deal. In anticipation of this, a regret-sensitive consumer will accept higher offers and therefore search shorter than someone who maximizes purely financial payoffs. Reduced depth of reasoning and impaired cognitive capacity due to time pressure can also lead to an inefficient search.

We use a pre-registered laboratory experiment in which we manipulate perceived urgency and regret to test our predictions. In the experiment, participants sequentially request price offers and incur a fixed search cost for every offer that they request (e.g., Schotter & Braunstein, 1981; Sonnemans, 1998). Thus, the participants themselves decide whether to continue the search and get another offer or to take the best stand-

[2] This chapter is published as Klimm et al. (2023). Time Pressure and Regret in Sequential Search. Journal of Economic Behavior and Organization 206, 406-424 in a slightly modified version.

ing offer. By manipulating whether or not information on post-purchase price realizations is available, we experimentally vary whether participants can feel regret about stopping the search too early. In the flight search example, this could mean that a consumer learns about prices on competing platforms after their purchase and regrets stopping too early if these prices are more favorable. Theoretically, the anticipation of experiencing regret when stopping too early prolongs the search. In addition, we analyze how perceived urgency affects search behavior and the role of regret on search behavior.

We find that anticipated regret has no effect on search behavior, while the experience of regret leads to systematic adjustments in search length. Urgency reduces decision times and perceived decision quality, but only very inexperienced decision-makers buy earlier under time pressure. With this, we contribute to policy-debates on pressure selling in dynamic market environments and highlight that consumer protection policies against sales tactics that rush consumers into making a decision may be especially helpful for inexperienced consumers. In addition, we highlight that committing to a reservation price prior to search is a viable strategy to overcome potential biases.

Chapter II is based on the idea that we often *like to be liked*.[3] In many settings, we prefer to interact with those who also want to interact with us (e.g., Avery & Levin, 2010; R. M. Montoya & Horton, 2012; Antler, 2019). Job applicants want to be the first-choice candidate, schools want to attract students who want them most, and singles want to go on a date with someone who is genuinely interested in them. We say that individuals who prefer to be matched with a partner who wants to be matched with them have *reciprocal preferences*.

Reciprocal preferences are particularly relevant in matching markets, where participants not only choose their partner, but must be chosen as well. While standard matching theory assumes that agents do not care about the preferences of their potential partners, there is evidence that participants want to know others' preferences before choosing their partner, to take into account who wants them most. Many markets even have mechanisms that allow participants to signal their preferences. Theoretically, reciprocal preferences can lead to situations where matching markets do not function as intended (Opitz & Schwaiger, 2023b). It is therefore crucial to understand the empirical relevance of reciprocal preferences for matching markets.

[3]A pre-print working paper version is available as Opitz and Schwaiger (2023). Everyone likes to be liked: Experimental Evidence from Matching Markets. CRC TRR 190 Discussion Paper, No. 366.

We identify reciprocal preferences and their implications for matching markets through a laboratory experiment. The experimental setting allows us to observe participants' preferences under different information sets. In the experiment, participants form two-person teams for a cooperative task, a Public Goods Game (PGG), through a centralized matching mechanism. Based on responses to a personality questionnaire, participants indicate with whom they would like to play the PGG. In one experimental condition, some participants learn with whom they are tentatively matched and how their potential partners rank them. In the other experimental condition, participants never learn how their potential partners rank them and only see with whom they are tentatively matched. Hence, we directly test whether agents' preferences are sensitive to information about others' preferences. We hypothesize that participants *like to be liked*, so that they prefer a partner who ranks them favorably. Therefore, participants would change their preference order after learning how others ranked them, leading to instability in the matching market. Thus, the experiment allows us to test whether individuals indeed prefer to interact with those who want to interact with them, to investigate the underlying mechanisms, and to demonstrate that this is a rationale for why matching markets can be unstable.

We provide evidence that reciprocal preferences exist, significantly decrease stability in matching markets, and are driven by both belief-based and preference-based motives. Participants expect partners who want to be matched with them to be more cooperative, and are more altruistic themselves. On average, knowing whether one is liked leads to more cooperation and higher profits. By understanding why matching markets may fail to achieve their objective, our results help design matching markets more effectively. Moreover, we contribute to the understanding of social preferences and social proximity, and point to broader organizational implications for team formation and teamwork.

Chapter III takes the importance of teamwork for organizations and the important role of interpersonal preferences for teamwork as a starting point. It builds on the findings of Chapter II that individuals often prefer to interact with those who also want to interact with them. I examine how interpersonal preferences affect non-routine teamwork and hypothesize that being liked may be important for effective collaboration. Understanding the role of interpersonal preferences for the success of teams is important because teams play an integral part in solving complex non-routine tasks within modern organizations (Autor & Price, 2013; Ichniowski & Shaw, 2013). These tasks require teams to assign responsibilities, work together, communicate effectively, share information or make joint decisions to solve problems.

I causally investigate whether teams perform better in complex problem solving when members like each other through a laboratory experiment. The experiment consists of a team formation process and a non-routine team task. During the team formation process, participants indicate with whom they want to interact in a payoff-irrelevant situation. That is, participants rank each other according to the desirability of interacting with each other. This is how I operationalize *liking*. These interpersonal preferences are based on self-reported questionnaire information of the potential partners. The incentivized non-routine task is played with one randomly matched partner. I analyze behavior under two information structures, similar to Chapter II. In one experimental condition, participants never know how much their partner likes them. In the other experimental condition, participants receive this information before the non-routine team task. I investigate whether performance in the non-routine task differs depending on how much team partners like each other, and I analyze different mechanisms of how this may affect team performance.

I find that interpersonal preferences matter for performance in complex problem solving. While teams in which partners like each other perform similarly to those in which partners dislike each other, teams in which one partner likes the other more (*dissimilar liking*) perform best. This is driven by changes in collaborative behavior upon learning the preferences of the partner, not by complementarities in skills of partners who display *dissimilar liking*. I provide suggestive evidence that one of the channels is different communication patterns. Participants do not expect to be more successful in teams with *dissimilar liking*. Rather, they expect to be most successful in teams where partners like each other, which is consistent with the findings in Chapter II.

Beyond understanding the role of interpersonal preferences in teamwork, these findings also have important implications for team formation. Because participants believe they will be more successful with those they like, self-selection can lead to inefficient outcomes. While in stylized one-shot interactions like those in Chapter II such beliefs map closely into actions (and payoffs), the determinants of success in the collaborative problem solving environment of this study are more complex. I illustrate that one key to success is effective communication, which should not be equated with the amount of communication. Teams that self-selected based on their interpersonal preferences may communicate more, but not necessarily more effectively. This raises important organizational questions about how much autonomy individuals should have in forming teams and how to most efficiently form teams to solve complex problems.

Chapter IV aims at unlocking the potential of entrepreneurial ideas and transforming start-ups into successful businesses by studying the role of gendered access to finance. While entrepreneurship is key for economic development, many businesses in the developing world stay small forever, rarely growing beyond subsistence size (La Porta & Shleifer, 2008; Hsieh & Olken, 2014). We address one major constraint to successful entrepreneurship: the lack of financial resources to grow their businesses. This constraint is particularly relevant for women who are less likely to have the necessary funding to start a business, face challenges in attracting external equity, and have more pronounced constraints on debt financing (e.g., OECD, 2017; Hebert, 2020; Brock & De Haas, 2023). In order to design targeted policies to close the gender gap in access to finance, it is necessary to understand the extent to which it is driven by supply-side factors, the role of gender bias on the supply side, and the underlying mechanisms of gender bias.

We focus on such mechanisms for gender bias on the supply side of access to finance. Specifically, we analyze whether loan officers' assessment of a start-up's future business performance depends on the entrepreneurs' gender and the team composition of the start-up. Our pre-registered lab-in-the-field experiment combines real-life data on start-up business performance with experimental measures on the assessment of start-up business ideas by loan officers in Uganda. By randomly manipulating the gender and team composition of the entrepreneurs on the business ideas, we are able to draw conclusions about whether and why women's ideas are evaluated differently. First, our design enables us to distinguish gender differences in the evaluation of the business idea itself from the evaluation of the entrepreneurs' implementation challenges and capabilities. Second, we can understand whether the formation of entrepreneurial teams can overcome biased evaluations.

We find a sizable gender bias for businesses by individual entrepreneurs, but no similar gender bias for teams of two entrepreneurs. For individuals, loan officers invest less in businesses by female entrepreneurs, are less likely to select a female entrepreneur's pitch deck as the best business among those they evaluated, and consider a woman's business to be significantly more likely to fail than a man's business. Our results suggest that this gender bias stems from a different assessment of women's entrepreneurial ability or a different assessment of the external constraints to their business implementation. In contrast, we do not observe a similar gender bias in the evaluation of teams of two entrepreneurs. Loan officers do not invest differently in businesses when they have a female founder or implementer on the entrepreneurial team. They do not perceive the ideas from female teams to be of lower quality, nor do they believe that female teams struggle more with the implementation of their business. We find that the contrasting results for individuals and teams are not driven by relative unfamiliarity with entrepreneurial teams or by different preferences and beliefs of loan officers about the two types of businesses.

The contrasting results for individual entrepreneurs and teams have important implications for understanding the financial barriers to female entrepreneurship in low-income countries. First, individual female entrepreneurs suffer from low access to finance because loan officers perceive them as having more difficulties in implementing their business. Second, we show that the formation of entrepreneurial teams can overcome biased perceptions, possibly by signaling higher growth aspirations. Therefore, we add nuance to the discussion of the role of gender in access to finance. We highlight that forming entrepreneurial teams can have second-order benefits for female entrepreneurs by changing perceptions about the business, which in turn can enable access to finance.

By studying generalizable mechanisms, this dissertation provides insights into the behavioral foundations of different stages of innovation processes, science, and entrepreneurship more broadly. Search is of much broader importance than just in consumer behavior. Search processes are equally relevant when thinking about the inputs into knowledge production as well as the output of science. Finding the right literature, screening for prior art, or comparing the available research tools can all be understood as important search processes in knowledge production. Similarly, the question of when to write up a project, or when to terminate it can be seen through the lens of a search problem. It is also natural to think about matching and collaboration in innovation processes. This is true in companies that increasingly rely on teamwork, but it is also true in research and in the formation of entrepreneurial teams. In all of these settings, it is important to understand the mechanisms of team formation that lead to successful cooperation and collaboration – both in terms of short-term output and persistence of these teams. Finally, the issue of bias in the evaluation of women's ideas applies more broadly than in the context of entrepreneurial finance in low-income countries. Similar biases may play an important role in peer review of papers and grants, mentoring programs, and funding decisions more generally. Addressing gender biased evaluations may be particularly valuable in these settings as well. Also the extent to which individual female contributions are evaluated differently compared to female teams is important in the context of co-authorship and other collaborations.

In summary, this dissertation offers new insights into the behavioral foundations of consumer choice, teamwork, and entrepreneurial finance by illustrating the role of preferences and constraints in decision-making. I use experimental techniques to address policy-relevant questions that have been difficult to answer using traditional econometric strategies. In doing so, I identify concrete mechanisms that can be turned into solutions to practical problems – possibly even beyond the core research questions I address in this thesis.

1

Time Pressure and Regret in Sequential Search

1.1 Introduction

Perceived urgency and regret are common in many markets. For instance, in many goods and service markets, sellers pressure buyers searching for the best price with time-limited offers and emphasize potential regret about forgone purchasing opportunities (Sugden, Wang, & Zizzo, 2019). In labor markets, job seekers face deadlines and anticipate (or experience) regret when they reject or accept offers. In financial markets, investors facing rapid price changes may regret forgone selling opportunities

*This chapter is based on joint work with Felix Klimm, Martin Kocher, and Simeon Schudy.

when holding onto badly performing assets (Strack & Viefers, 2021).[1] It is thus important to understand to what extent perceived urgency and regret may affect individual choice in dynamic market environments, and whether their combination aggravates or alleviates potential biases in decision making.

Our study investigates the effects of perceived urgency and regret in a pre-registered, theory-based laboratory experiment.[2] Many of the above-mentioned examples for the relevance of urgency and potential regret reflect a search process that can be represented by an optimal stopping problem. In optimal stopping problems, a decision-maker observes a sequence of realizations of some stochastic process and, after observing a realization, decides on whether or not to take an action. For example, buyers may learn about price offers for a flight and then decide on whether to continue searching for a better realization (e.g., by looking at other platforms or waiting another day) or they may stop searching and immediately buy the item for the best available price.[3]

By trading off the best current price with potentially better future prices at higher search costs, decision-makers may experience regret of two types. First, if it turns out that decision-makers could have saved unnecessary search costs, they may regret not having stopped searching earlier (which is often referred to as *inaction regret*). Second, when deciding on whether or not to accept the currently best available price, decision-makers may anticipate that better price realizations can become available after purchase, and thus may anticipate regret from not having searched for longer (i.e., if they observe price realizations after purchase, which is often referred to as anticipated *action regret*).

While an expected utility maximizer is assumed to calculate the optimal search length given her knowledge about the underlying stochastic process and given search costs, perceived urgency may render full optimization unlikely. Time-pressured individuals may rely more on intuitive rather than deliberative decision making (Epstein, 1994; Kahneman, 2003, 2011), use heuristics to a greater extent (Gigerenzer & Todd, 1999), or forgo a thorough and in-depth processing of available information (Kruglanski &

[1] In addition, urgency and regret are prevalent in auctions. For instance, in first-price auctions, bidders may anticipate or experience regret when paying too much (relative to the second-highest bid) when winning, or when bidding too little and thus missing an opportunity to win the auction at a favorable price (Engelbrecht-Wiggans & Katok, 2008).

[2] Pre-registration at: AEA RCT Registry; AEARCTR-0004065.

[3] The best available price relates either to the current price offer (optimal stopping with no recall) or the best price among the current and past price offers that the buyer has observed (optimal stopping with recall).

Freund, 1983).[4] Furthermore, perceived urgency may not only result in lower levels of choice accuracy, but may also alleviate anticipated *action regret* because anticipation of regret is less salient when there is (or appears to be) limited time to deliberate.[5]

Our experiment disentangles these channels in a parsimonious dynamic decision-making environment that allows us to identify the role of regret, perceived urgency, and their interaction. Participants in the experiment buy one unit of a product and maximize their payoff by purchasing the item at a low price without searching for too long. They can sequentially request additional price offers and incur a fixed search cost for every offer that they request (see also Schotter & Braunstein, 1981; Hey, 1987; Cox & Oaxaca, 1989; Kogut, 1990; Sonnemans, 1998). In other words, the participants themselves decide to continue the search for another round or to take the best standing offer. They know the distribution from which offers are drawn and that all previously observed offers are attainable (i.e., we employ optimal stopping with recall). Consequently, expected profit maximization is characterized by adherence to a constant reservation price strategy (Lippman & McCall, 1976). Expected payoff-maximizing individuals search until an offer at or below their reservation price is observed and they then buy the item at that price.

Two deviations from the constant reservation price strategy are commonly observed in search environments, in which buyers do not receive post-purchase information on prices: early stopping and the recall of previously rejected prices. Regardless of the context, previous studies show that participants request fewer offers than theoretically predicted (e.g., Hey, 1987; Cox & Oaxaca, 1989; Sonnemans, 1998; Houser & Winter, 2004; Einav, 2005) and they often make use of the recall option e.g., Schotter and Braunstein, 1981; Hey, 1987; Kogut, 1990; Houser and Winter, 2004; Ibanez, Czermak, and Sutter, 2009; Schunk, 2009; Schunk and Winter, 2009, which is in line with the idea of anticipated *inaction regret*. Indeed, expanding a standard sequential search model (Lippman & McCall, 1976) by regret aversion predicts both of these commonly observed patterns of behavior (see Appendix A.1 for more detail). Consequently, we designed our experiment to ensure that we can empirically assess the relevance of *regret*. By manipulating whether or not information on post-purchase price realizations is available, we exogenously vary whether anticipated *action regret* can prolong search, countervailing the potential effects of *inaction regret*. Further, we

[4]As has been shown, for instance, in the context of risk-taking and loss aversion (see e.g., Ben-Zur & Breznitz, 1981; Kocher, Pahlke, & Trautmann, 2013; Kirchler et al., 2017).

[5]This idea is in line with the finding that, when explaining individuals' behavior with drift-diffusion models, time-pressure reduces barrier height to speed up choices (Milosavljevic et al., 2010).

employ random variation in feedback to study the role of experienced *action regret*. As buyers are also often pressured time-wise, we further study how perceived urgency alters search behavior and the role of regret. We implement a 2x2 between-subjects design with high or low perceived urgency that avoids potential selection bias due to time pressure, and vary search costs (within-subjects) to analyze the extent to which participants understand the general logic of the reservation-price strategy.

Our empirical results confirm stylized facts from previous experiments, as in all treatments, participants search on average too little (as compared to the expected payoff-maximizing strategy), make use of the recall option, and search longer with lower search costs and more experience. In our main analyses, we study the causal role of perceived urgency, regret, and their interaction for search behavior. We find that perceived urgency reduces decision times and perceived decision quality but does not change search length in general. However, in the very first search task, time pressure does affect search length and reduces payoffs substantially. Anticipated *action regret* (i.e., anticipating regret from stopping too early) does not increase search length, while experienced regret, both *action* and *inaction regret*, leads to systematic adjustments in search length. Learning that one has stopped searching too early, leads to longer search in the subsequent task while searching for too long reduces search length. These adjustments do not increase payoffs substantially, as some participants over-adjust their search length. Finally, perceived urgency does not substantially alter the observed role of regret.[6] In addition to our main analysis, our study highlights the need for strategies consumers may employ to protect themselves from searching sub-optimally. Thus, we also discuss commitment to reservation prices as a simple strategy that may circumvent inefficient search and provide empirical evidence showing that such commitment can indeed improve the optimality of search and results in larger payoffs.

The rest of this manuscript is organized as follows. Section 1.2 discusses the three-fold contribution of our approach (i.e., understanding the role of time-pressure, regret, and their potential interaction in sequential search tasks) relative to the existing literature. In Section 1.3, we explain the experimental design. In Section 1.4, we specify theory-based hypotheses, which we test in our main empirical analyses in Section 1.5. In Section 1.6, we discuss our findings and their robustness. Section 1.7 concludes.

[6] Importantly, our experiment allows to identify economically relevant effect sizes (i.e., larger than 0.20 standard deviations, for more details, see also Section 1.7.

1.2 Related Literature

Our search design builds on classical search experiments (e.g., Schotter & Braunstein, 1981; Sonnemans, 1998; Houser & Winter, 2004) which revealed two commonly observed anomalies in sequential search problems: early stopping and recall. Our experimental treatment variations complement and advance earlier experimental findings on active sequential search under conditions with or without perceived urgency as well as with or without post-purchase price information.

Our analyses on perceived urgency extends earlier experimental findings on time pressure in sequential search environments that excluded post-purchase price information. Ibanez, Czermak, and Sutter (2009) document inefficiently short search patterns for inexperienced decision-makers under mild time constraints (without post-purchase feedback). Through making deliberation more costly in our design, we confirm that time pressure substantially reduces payoffs with inexperienced decision-makers both with and without post-purchase price information. These causal experimental findings are also consistent with correlational evidence from the field, which shows that urgency due to being close to a purchasing deadline is associated with decreased search in an environment with price uncertainty (Lemieux & Peterson, 2011).

Regarding the study of anticipated *action regret*, our approach links to the literature that has studied the effects of post-purchase information on search behavior. Sugden, Wang, and Zizzo (2019) study whether time-limited offers are chosen more often without post-purchase information, finding no evidence of regret effects. In contrast, we focus on how feedback structures and perceived time pressure affect the number of requested (ex-ante identical) offers. In line with the findings of Sugden, Wang, and Zizzo (2019), we provide robust evidence on the limited role of anticipated *action regret* for search length when decision-makers actively incur search cost to receive additional offers. Our findings further complement important recent evidence on the search-enhancing effect of anticipated *action regret* when decision makers search through repeatedly stating reservation prices and post-purchase information only includes (potentially) better offers (Jhunjhunwala, 2021). Relating to this work, we provide evidence from additional experimental treatments (see Section 1.6.3) which underscores the critical role of the nature of post-purchase information which may generate behavioral changes through anticipated regret.

More generally, our results regarding anticipated and experienced regret relate to the broader literature on optimal stopping problems. Strack and Viefers (2021) demonstrate regret sensitivity in an asset-selling task where new offers are automatically updated at no monetary cost and decision-makers have no recall option. To distinguish the behavior of a regret agent from an expected payoff-maximizer, the empirical analysis of Strack and Viefers (2021) relies on random choice behavior. In their analysis, they assess an agent's sensitivity to feelings of *inaction regret* after having continued the search when it was optimal to stop.[7] Our analyses also link to work by Fioretti, Vostroknutov, and Coricelli (2022), who vary (within-subject) post-purchase information in a setting akin to Strack and Viefers (2021) and find –consistent with our theoretical predictions– that participants stop later when they may anticipate *action regret*.[8] While these studies focus on situations in which new prices arrive automatically and no recall option exists, our approach involves an active, costly choice for new price requests and allows for recall. These changes may render the role of regret less salient in our setting. On the other hand, avoiding *action regret* may be perceived as less costly in Fioretti, Vostroknutov, and Coricelli (2022). The stochastic mean-reverting process that determines the prices in Fioretti, Vostroknutov, and Coricelli (2022) leads to a multimodal distribution of prices (and payoffs) over time. Thus, it becomes likely that participants encounter similar payoffs in the future, even when not selling early on. As the cost of delaying the purchase in early periods becomes less costly, participants may stop later and at the same time achieve similar payoffs while reducing the probability of *action regret*.

Finally, our setup allows us to study the role of experienced regret which may induce learning across time (see e.g., Sonnemans, 1998; Cooke, Meyvis, & Schwartz, 2001; Einav, 2005; Oprea, Friedman, & Anderson, 2009). Sonnemans (1998) (Experiment 2) shows that participants change their reservation prices after learning that they searched too long. Similarly, participants converge faster to an optimal reservation price in a search task with pre-commitment when receiving post-purchase feedback (Einav, 2005). Oprea, Friedman, and Anderson (2009) provide post-purchase price realizations in all treatments of an investment task and observe that regret associated with stopping decisions in past tasks leads participants to reconsider their strategy in future tasks. This is in line with findings on the learning-enhancing effect of regret

[7]Our theoretical predictions are in line with those of Strack and Viefers (2021) for *optimal stopping*. However, their information structure does not allow them to analytically discriminate between a decision-maker with regret aversion and an expected utility decision-maker when analyzing *optimal stopping*.

[8]Note that contrary to classical experimental search tasks, the environment of Fioretti, Vostroknutov, and Coricelli (2022) already leads to longer search than theoretically predicted in the condition without post-purchase feedback, while the classical anomaly in search tasks goes in the opposite direction compared to the rational benchmark.

through priming (Reb, 2008; Reb & Connolly, 2009). Our results complement this line of research. In general, we find that the fraction of searches that are too long remains constant across time while the fraction of searches that are too short decreases within the first half of the experiment, thereby reducing inefficiencies to some extent. Experienced (action and inaction) regret alters search length in our setting systematically. In particular, participants in the treatment condition with post-purchase information increase (decrease) search length after experiencing action (inaction) regret. However, such learning from experienced regret does not translate into higher levels of efficiency, presumably because participants face different search costs and prices across search tasks, rendering profitable adjustments more complex. Finally, experiencing *action regret* from searching to little does not reinforce anticipated regret. That is, differences in search lengths across feedback conditions do not substantially change across the 10 search tasks.

1.3 Experimental Design

The main part of the preregistered experiment consists of 10 standard sequential search tasks and two additional search tasks with pre-commitment on a reservation price (see also Einav, 2005). For the 10 sequential tasks, we vary perceived urgency by inducing high or low time pressure (*High-TP*, *Low-TP*) and whether participants can anticipate *inaction regret* by providing feedback on post-purchase price offers (*Info*, *No-Info*) in a 2x2 between-subject design, while holding all other aspects of the decision environment constant. After the main part of the experiment, we elicit incentivized measures for the participants' expected relative performance, risk attitudes, and loss attitudes. Furthermore, we elicit a subjective, non-incentivized measure of decision quality relative to participants in the alternative time-pressure condition, and we collect information on socio-demographic characteristics in a short post-experimental questionnaire. At the end of the experiment, one of the 12 search tasks is randomly drawn to be payoff relevant. Figure 1.1 summarizes the experimental procedures, showing the different parts of the experiment. To avoid unwanted effects of anticipating the content of subsequent parts, we inform participants only at the beginning of each part about its content. Further, participants of the subject pool are aware that they receive a flat payment of 6 Euro and that they can make losses during some parts of the experiment which will be compensated by the 6 Euro flat payment and potential earnings from other experimental parts. For example, given the nature of the search task in our experiment, participants could encounter losses in Tasks 1-12, if they decided to pay a price higher than their valuation or when searching too long and thus incurring search costs larger than the gains from trade.[9]

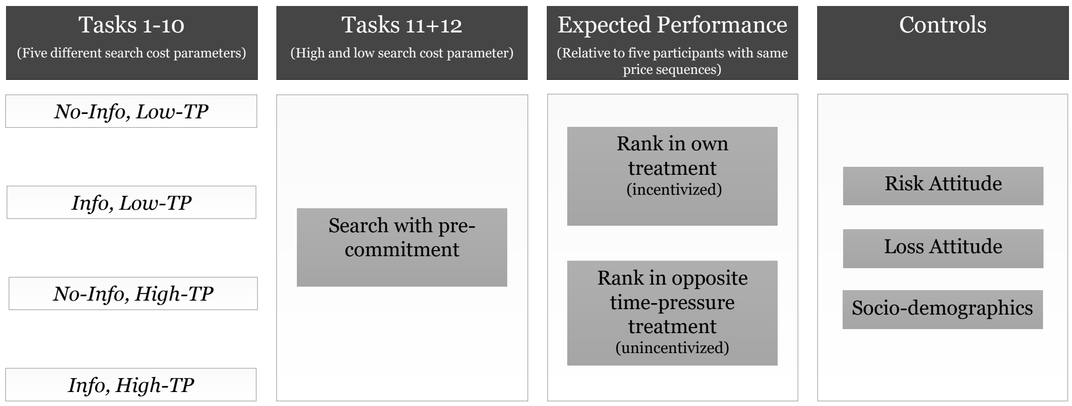


Figure 1.1: Experimental Design

1.3.1 Sequential Search Tasks

Participants decide in 10 sequential search tasks whether to buy a fictitious product at the best price observed so far (i.e., optimal stopping with recall).[10] The participants' induced value for the good is $v = 50$ and stays the same across all search tasks. At the beginning of each search task, participants see a first price offer at which they can buy and they then decide whether or not to accept the price or ask for an additional offer. Each additional offer comes at a fixed cost c, which stays constant within each of the ten search tasks (but varies across tasks) such that participants are aware about their search cost when deciding upon an additional price request. Price offers are drawn from the known uniform distribution $\{1, 2, ..., 100\}$.[11] We inform the participants that they are free to request new offers as long as there is a possibility to achieve a positive payoff given search costs. This renders the search process finite (because participants can request at most 24 additional offers before making a loss for sure given our parameter values, although we do not state the exact number of possible requests to participants).[12] After purchasing the product the current search task ends

[9] All participants received positive payoffs and this procedure does not alter our theoretical predictions.

[10] With perfect recall, previous prices serve as a form of insurance against unsuccessful draws. This reduces the role of risk attitudes on search behavior, allowing us to neatly examine the role of regret.

[11] We thereby rely on the parametrization of Sonnemans (1998).

[12] Only in 0.26 percent of all decisions were 24 additional prices requested (by a total of 4 out of 191 participants). In these cases, the computer automatically bought the product at the best standing price.

and participants proceeded with the next search task.

1.3.2 Price Sequences and Search Costs

Price sequences were determined randomly in the first two sessions. To keep sequences constant across treatment conditions, the same randomly drawn sequences are used in later sessions. We form within-treatment clusters of six participants who received the same 10 randomly drawn price sequences for the 10 search tasks. Hence, our design allows for a between-subject but within-sequence comparison. Each search task contains eight independent price sequences (because we have 48 participants per treatment and a cluster size of six), and thus the 10 tasks include 80 independently drawn price sequences. To ensure that perceived urgency can affect search behavior also in later tasks, we vary the theoretically optimal reservation price strategy by altering search costs between the tasks. We use five different values for the search cost $c \in \{2, 2.5, 3, 3.5, 4\}$. Each parameter value occurs twice and the order in which these parameters appear is randomly determined but held constant for each price sequence and announced for each task as it starts.

1.3.3 Experimental Treatments

Time pressure

We exogenously vary perceived urgency by limiting the amount of time that an individual can spend on each search step (i.e., deciding about buying the product vs. requesting another offer). Instead of resorting to strict time constraints see, e.g., Ibanez, Czermak, and Sutter, 2009; Sugden, Wang, and Zizzo, 2019, we induce perceived urgency by making longer deliberation more costly. In our high time pressure treatment *High-TP*, participants incur a monetary punishment (1 Taler = 1 unit of the experimental currency) if they fail to accept or ask for a new offer within 4 seconds (and the computer deduces 1 additional Taler every 4 seconds if no decision is made). In our low time pressure treatment *Low-TP*, we set the time limit to reflect on each offer to 60 seconds (i.e., the computer deduces 1 Taler every 60 seconds if no decision is made). This procedure avoids unwanted selection effects of drop-outs without a deliberate decision (see e.g., Kocher et al., 2019), which allows us to impose time pressure without forcing participants to accept a default (or random) decision after the time ran out and excludes participants from intentionally avoid submitting a choice at all.

Anticipated Regret

Orthogonal to the variation in perceived urgency, we vary the feedback after the purchase decision has been made; and thereby, whether decision-makers can anticipate *action regret* from stopping too early. In treatment *No-Info* participants are informed that they see only those prices that they actively requested until they purchase the product. In treatment *Info*, the participants are informed that they will see additional price offers for which they could have bought the product, after purchasing it. We randomly determine the number of displayed offers $k \leq n$ where $n = 25 - OfferNumber_{accepted}$, such that (for example) a participant who decides to buy after seeing five offers can see between 1 and 20 additional prices. This design feature renders learning about the maximum possible search length similar in both treatments. By varying the availability of post-purchase information, we thus exogenously vary whether or not the participants can anticipate action regret from buying too early (see also Zeelenberg, 1999; Sugden, Wang, & Zizzo, 2019; Jhunjhunwala, 2021; Fioretti, Vostroknutov, & Coricelli, 2022). This anticipation can be reinforced, when experiencing *action regret* in *Info* in previous tasks. Because we randomize the number of additional prizes displayed, we vary whether participants experience regret given the same search behavior and price sequence. This allows us to analyze the effect of experienced action regret both within the *Info* treatment and across treatments, and disentangles potential effects of simply seeing additional prizes (e.g., by familiarizing oneself with the random process of price draws) as compared to experiencing regret due observing particularly attractive prices.

1.3.4 Search Tasks with Pre-Commitment

After the 10 sequential search tasks, we confronted all of the participants with two additional search tasks that allow for pre-commitment. In these tasks, the participants pre-specify a price at or below they are willing to buy the good and face no time constraint in that choice. The computer then draws offers until the threshold is reached or undercut. Irrespective of the treatment, the participants have been assigned in the 10 sequential search tasks described earlier, we provide no post-purchase information on additional prices in the tasks with pre-commitment. Thus, the feedback structure rules out anticipated *(action) regret*, and pre-commitment avoids *experiencing (inaction)* regret during the task (as well as the use of the recall option). Search with pre-commitment and without time pressure may therefore counteract potential biases through regret and time pressure. One of the two search tasks involves low search costs ($c_{min} = 2$) and the other involves high search costs ($c_{max} = 4$). This variation also allows us to cleanly test for the participants' responsiveness to the search costs.

1.3.5 Evaluation of Own Performance

After the 12 search tasks, the participants have to guess their performance rank (1st to 6th) among those participants who saw the same price offers (i.e., in the within-session price sequence cluster). The subjects are incentivized by a monetary payment if their stated rank matches the actual decision quality (rank) and they receive no payment otherwise. In addition, the participants guess their rank in comparison to the participants who saw the same price sequences and were assigned to the same feedback (*Info* / *No-Info*) condition but to the other time pressure condition. This second, unincentivized measure allows us to study whether participants consider the exogenous increase in perceived urgency to be a less (or more) favorable decision environment.

1.3.6 Control Variables

Given that risk aversion may theoretically shorten search length (empirically, it does not seem to do so, see also Sonnemans, 1998; Schunk & Winter, 2009), we elicit an incentivized proxy for risk attitudes, using the approach by Holt and Laury (2002). We also measure the participants' loss attitudes following the incentive-compatible procedure by Gächter, Johnson, and Herrmann (2022), as suboptimally short search durations may be driven by loss aversion see e.g., Schunk, 2009. Finally, the participants complete a standard socio-demographic questionnaire (including gender, age as well as their final math grade in high school).

1.3.7 Procedures

The experiment was conducted at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) in July and August 2019. In total, 192 participants took part in the experiment.[13] We ran eight sessions (with 24 participants each, two sessions per treatment). The participants were recruited using the online system ORSEE (Greiner, 2015), and we restricted participation to students without experience in sequential search tasks. The experiment was programmed with the software z-Tree (Fischbacher, 2007). On average, participants earned 20 EUR (including a show-up fee of 6 EUR), and the experiment lasted around 60 minutes. Each session was supervised by the same experimenters.

1.4 Predictions

Our main hypotheses concern search behavior; that is, they are directed at differences in the number of requested offers within and across treatment conditions. We also investigate how the number of requested offers corresponds to (ex-ante) efficiency and actual payoffs.

1.4.1 Regret

Our predictions on the role of regret are based on a theoretical model (see Appendix A.1) which incorporates regret aversion in sequential search building on the formulations of Schunk (2009). This model, reconciles both frequently observed anomalies in empirical search settings without post-purchase information. It predicts that regret-sensitive participants have a higher reservation price (i.e., they request fewer offers) compared to the rational benchmark as they may suffer from *inaction regret* (i.e. from not stopping early enough). The model is also consistent with moderate rates of recall within a task due to *inaction regret*. We specify this prediction in Hypothesis 1:

Hypothesis 1. *In treatment No-Info, regret aversion leads to fewer requested offers when compared to the risk-neutral, regret-free benchmark and it also allows for the use of the recall option.*

The model further predicts that participants request more offers when they know that

[13]We excluded one participant from the analysis because their search behavior was unresponsive to prices and incentives from task 3 onwards; that is, the participant requested the maximum amount of offers in 8 out of 10 tasks, even when already having encountered extremely favorable offers. Additionally, the decision times of this participant were the fastest across all participants in *Low-TP*. The analyses including this participant are qualitatively the same and can be found in Appendix A.4.1.

post-purchase information will be shown (*Info* vs. *No-Info*) because the participants can only regret having stopped too early when learning post-purchase price information. Anticipating this *action regret* theoretically prolongs search lengths. We summarize this prediction in Hypothesis 2:

Hypothesis 2. *With anticipated (action) regret, the number of requested offers is lower in treatment No-Info than in treatment Info.*

We additionally hypothesize that experiencing regret reinforces anticipated regret, induces directional learning, and systematically influences search behavior in subsequent tasks. Decisions in repeated search tasks may reflect the experience of regret in the previous task, translating into higher awareness and sensitivity to anticipated regret. For Tasks 2 to 10, we specify below one hypothesis for *inaction regret* (i.e., not stopping early enough) that can be present in both information structures and one hypothesis for *action regret* (i.e., having stopped too early) that can only arise under *Info*. We hypothesize that experiencing *inaction regret* leads to a lower number of requested offers in the subsequent search task, whereas we expect experiencing *action regret* to lead to a higher number of requested offers in the subsequent search task.

Hypothesis 3. *The experience of inaction regret (having searched too much) in task k leads to a lower number of requested offers in task $k+1$ in treatments Info and No-Info.*

Hypothesis 4. *The experience of action regret (having searched too little) in task k leads to a higher number of requested offers in tasks $k+1$ in treatment Info.*

Note that empirically testing Hypothesis 2 across all tasks combines the effect of anticipated and experienced regret. In Tasks 2-10, the participants may already have experienced regret in previous tasks, which can directly enhance learning or reinforce the anticipation of regret. To isolate the effect of anticipated regret, we additionally compare search lengths across treatments (*Info* and *No-Info*) in the very first search task participants encounter. Because the participants did not experience regret before this task, the differences between both treatments can be attributed entirely to the anticipation of seeing additional (potentially more favorable) price realizations.

1.4.2 Time Pressure

Perceived urgency has been found to reduce the depth of reasoning and alter information processing (Payne, Bettman, & Luce, 1996; Kocher & Sutter, 2006). Altering participants' optimization process, perceived urgency may thus result in shorter or longer search length. The observation that sellers use practices that create a sense of urgency suggests a reduction in search length as higher accepted prices benefit sellers. Participants may also tend to accept current offers more frequently when they perceive pressure and thus consider the *High-TP* decision environment to be aversive. At the same time, time pressure may impair the availability of cognitive resources and thus render the consideration of additional psychological factors less likely. If these are the reason for (inefficiently) short search, time pressure may increase search length. Further, if participants rely increasingly on decision heuristics under time pressure (e.g., Gigerenzer & Todd, 1999; Finucane et al., 2000), search length may increase or decrease (depending on the decision heuristic). Because a priori both longer or shorter search is possible and any specific modeling choice seems somewhat arbitrary, the direction of impact remains an empirical question. Consequently, we do not specify a directed hypothesis and instead we formulate the null hypothesis that limiting the time to reflect on an offer does not affect search length.

Hypothesis 5. *The number of requested offers does not differ between treatments High-TP and Low-TP.*

1.4.3 Potential Interaction of Time Pressure and Regret

Building on the idea that time pressure renders the consideration of additional psychological factors less likely (unless they are automatically invoked in the form of heuristics), a potential increase in search length due to the provision of post-purchase price information (i.e., due to the possibility to anticipate regret from requesting too few offers in *Info* and the lack thereof in *No-Info*) should be lower under time pressure. The lower availability of cognitive resources leads to regret being less relevant for the decision. We summarize this prediction in Hypothesis 6, which relies on the assumption that our theory-based prediction for anticipated regret (Hypothesis 2) is also observed empirically:

Hypothesis 6. *Anticipated regret impacts search length to a lesser extent in environments with high levels of perceived urgency.*

1.5 Main Results

1.5.1 Search Behavior without Feedback

As outlined above, in this sequential problem, the optimal strategy for a payoff-maximizing regret-free and risk neutral agent is a constant reservation price strategy (see Lippman & McCall, 1976). That is, conditional on search costs, agents derive a cutoff value for the price below which they will buy the good (see also Appendix A.1).[14] Given search costs and realizations of prices in the 10 sequential tasks, this cut-off value translates into an (ex-ante) optimal search length of 4.56 offers in our setting.

In the experiment, however, we observe substantially shorter search lengths (see Table 1.1). Relating to earlier literature, we first focus on the standard environment without information about future prices (and discuss potential treatment differences in Section 1.5.3). Participants stopped on average after seeing 3.83 offers when receiving no information on post-purchase prices (pooled across both time pressure conditions; $p < 0.001$, Wilcoxon signed-ranks test).[15] This result also holds when analyzing each time pressure condition individually (*No-Info/Low-TP*: $p < 0.001$, *No-Info/High-TP*: $p < 0.001$, Wilcoxon signed-ranks tests). The search length corresponds to an average accepted price of 16.59. Consequently, the participants also earned around 11 percent less than the expected payoff-maximizer would obtain ($p < 0.001$, Wilcoxon signed-ranks test). Furthermore, in a substantial fraction of searches (18.84 percent), the participants make use of the recall option (similar to rates in previous studies between 10-30 percent (e.g., Schotter & Braunstein, 1981; Kogut, 1990; Ibanez, Czermak, & Sutter, 2009)), and 78.95 percent of participants do so at least once in the experiment. Recall rates do not differ statistically significantly across time pressure conditions in the standard search environment (17.02 percent in *No-Info/Low-TP*, 20.63 percent in *No-Info/High-TP*; $p = 0.435$, MWU). Hence, we find strong evidence in support of Hypothesis 1:

Result 1. *Participants request significantly fewer offers in No-Info than the risk-neutral and regret-free benchmark predicts and use the recall option.*

[14] Depending on the search costs, the reservation price is between 20 and 29 for an expected payoff-maximizer given our parametrization.

[15] For the non-parametric tests, we form within-subject averages across the respective tasks so that we consider one data point per individual. All of the reported non-parametric tests in the analysis are two-sided hypothesis tests.

Table 1.1: Decriptive Statistics on Search Behavior

	Search Length				Accepted Price	n
	Mean	Min	Max	SD	Mean	
No-Info/Low-TP	3.82	1	25	3.11	15.75	470
No-Info/High-TP	3.85	1	24	3.24	17.42	480
Info/Low-TP	3.74	1	25	3.40	16.44	480
Info/High-TP	3.73	1	21	2.97	17.91	480

This table shows descriptive statistics on search behavior across the four treatments. *Mean*, *Min*, *Max*, *SD* denote the mean, the minimum, the maximum, and the standard deviation, respectively. *n* denotes the number of observations.

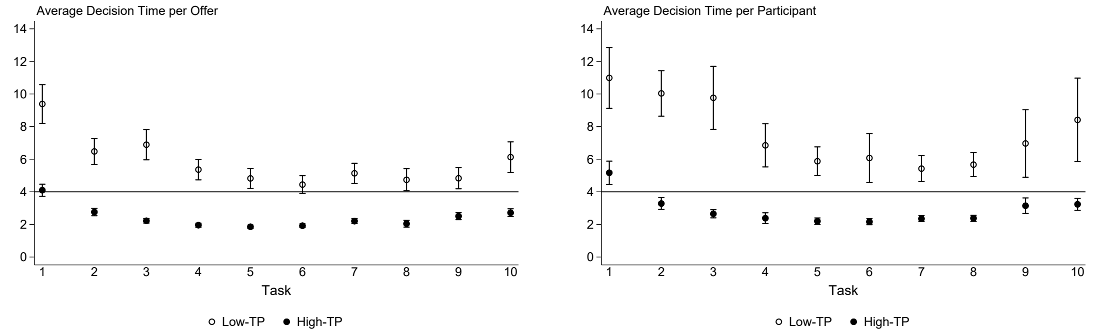
1.5.2 Manipulation of Perceived Urgency and Decision Times

Before we present the effects of regret and perceived urgency on search behavior, we briefly establish that our time-pressure intervention indeed resulted in shorter decision times. This is important because our *High-TP* condition deliberately avoids forcing the participants to decide within a strict time limit. Instead of implementing a deadline, the treatment makes slower decisions more costly by deducting 1 point for every 4 seconds that the decision-maker takes to reflect on a price offer. Hence, our treatment variation relies on the assumption that people perceive urgency, and therefore they mostly comply with the time limit.[16]

Our treatment manipulation regarding perceived urgency worked very well. Enforcing a time limit of 4 seconds would be binding in the vast majority of searches under *Low-TP*. Across all tasks, participants in *Low-TP* take 5.73s per decision; 44.64 percent of decisions in *Low-TP* take longer than 4 seconds. More importantly, Figure 1.2 and Table 1.2 highlight that decision times are substantially and statistically significantly shorter in *High-TP* than *Low-TP* in all sequential search tasks (pooled across both feedback conditions) [17] Furthermore, the fraction of tasks where all of the decisions were taken within 4 seconds is substantially lower in *Low-TP* when compared to *High-TP* (14.11 percent and 67.19 percent; $p < 0.001$, Mann-Whitney U test [MWU]). Hence, the participants indeed perceived urgency in *High-TP* and made faster decisions.

[16] Relative to the average earning in the search task, transgressing the limit once compares to a decrease in earnings of around 4 percent.

[17] Table A.1 corroborates that the decision times significantly decrease in both feedback conditions.



(a) *Average time per offer* **(b)** *Average time per participant across all offers*

Notes. The error bars indicate 95% confidence intervals.

Figure 1.2: Decision Times Across all Sequential Tasks for *Low-TP* and *High-TP*.

Table 1.2: Average Decision Times per Task across Time Pressure Conditions

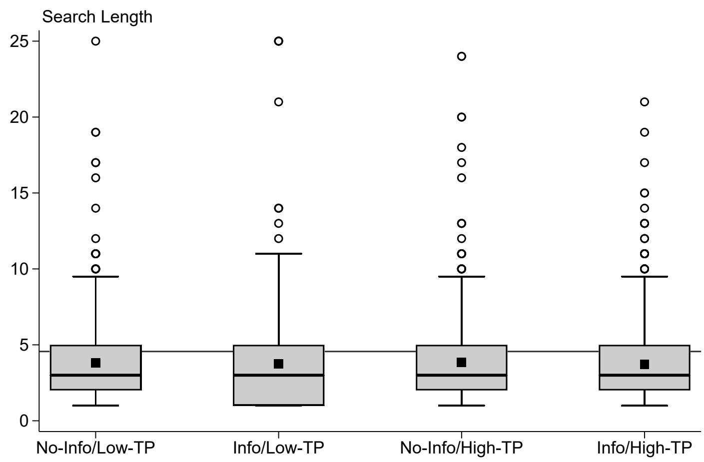
	per Offer		per Subject		
Task	*Low-TP*	*High-TP*	*Low-TP*	*High-TP*	p-value
1	9.39	4.10	10.99	5.17	<0.001
2	6.48	2.76	10.04	3.28	<0.001
3	6.89	2.22	9.77	2.65	<0.001
4	5.36	1.95	6.85	2.38	<0.001
5	4.82	1.86	5.87	2.20	<0.001
6	4.44	1.92	6.07	2.16	<0.001
7	5.13	2.20	5.42	2.35	<0.001
8	4.74	2.05	5.67	2.37	<0.001
9	4.83	2.50	6.97	3.15	<0.001
10	6.13	2.72	8.41	3.23	<0.001

The table shows the average decision times across the time pressure conditions. The p-values are based on non-parametric Mann-Whitney U tests (MWU) on whether the participants' average decision times per task in *Low-TP* and *High-TP* come from the same underlying distribution.

1.5.3 Search Length across Treatments

Related to Hypotheses 2 to 5, we compare search behavior across treatments. First, we consider all 10 search tasks jointly and analyze the average effect of time pressure. Then, we consider the joint effect of anticipated and experienced regret on search length. While it may be necessary to experience regret before adjusting behavior in subsequent decisions, a separate analysis of the very first task decision-makers encountered allows us to isolate the effect of anticipated (action) regret (see Section 1.5.5).[18]

Considering all 10 search tasks, the number of requested offers does not differ significantly across treatments. Neither do we observe a difference between *High-TP* and *Low-TP* (pooling in terms of Info, $p = 0.750$, MWU) nor between *No-Info* and *Info* (pooling in terms of time-pressure, $p = 0.646$, MWU). The same holds when comparing treatments individually instead of pooling them (see Table 1.1). Time pressure neither changes the number of requested offers without ($p = 0.941$, MWU) nor with feedback ($p = 0.575$, MWU); the feedback structure neither affects average search length without ($p = 0.451$, MWU) nor with time pressure ($p = 0.967$, MWU). Figure 1.3 illustrates that the average search length is below the (ex-ante) optimal benchmark of 4.56 offers (vertical line) and that the distributions of search lengths across treatments do not differ substantially.



Notes. The figure shows boxplots of search lengths across treatments and a vertical line that indicates the optimal (ex-ante) threshold of a risk-neutral regret-free participant. The length of the whiskers is 1.5 times the interquartile range. The mean search length of each treatment is indicated by a solid square. The vertical line within the box corresponds to the median.

Figure 1.3: Search Length across Treatments (Tasks 1-10).

We corroborate these findings in regression analyses (Table 1.3; Columns (1)-(3)). In Column (1), we assess the treatment effect, controlling for the number of tasks a decision-maker already completed. In Column (2), we add demographic controls,

[18]For completeness, we also provide a separate analysis of tasks 2-10. These results mirror the results when considering tasks 1-10 jointly and can be found in Appendix A.3.

Table 1.3: Search Length

	Number of offers					
	Task 1-10			Task 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments						
High-TP	.022	.072	.071	-.973**	-1.076***	-1.097***
	[-.461,.506]	[-.405,.549]	[-.378,.519]	[-1.737,-.208]	[-1.884,-.268]	[-1.796,-.397]
Info	-.086	-.045	-.059	-.327	-.211	-.214
	[-.571,.399]	[-.515,.425]	[-.474,.357]	[-1.188,.534]	[-1.103,.682]	[-.860,.432]
High-TP X Info	-.033	-.064	-.060	.910	.961	.968
	[-.704,.639]	[-.732,.603]	[-.663,.542]	[-.379,2.199]	[-.344,2.266]	[-.204,2.140]
# Tasks encountered	.079***	.079***	.079***			
	[.032,.125]	[.032,.125]	[.032,.125]			
Risk Aversion		-.036	-.067*		.002	-.080
		[-.117,.044]	[-.145,.011]		[-.176,.181]	[-.256,.096]
Loss Aversion		.017	.017		-.253*	-.233*
		[-.110,.145]	[-.100,.134]		[-.530,.024]	[-.470,.005]
Constant	3.391***	4.295***	4.764***	3.681***	5.747***	4.780***
	[2.988,3.793]	[3.301,5.289]	[3.549,5.979]	[3.081,4.281]	[3.414,8.081]	[2.263,7.297]
Socio-demographic controls	No	Yes	Yes	No	Yes	Yes
Price Sequence Group FE	No	No	Yes	No	No	Yes
Observations	1910	1910	1910	191	191	191

OLS Regressions.*** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. Values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, which represents the number of offers after which the participant stopped searching. Columns (1)-(3) display search behavior across tasks 1-10, columns (4)-(6) for the first task. Columns (1) and (4) show the effect of the treatments. Columns (2) and (5) add socio-demographic controls (gender, age, cognitive ability) elicited after all search tasks; columns (3) and (6) additionally include price sequence group fixed effects. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

as well as measures of risk and loss attitudes.[19] In Column (3), we add fixed effects for the price sequence cluster. In all of the specifications, point estimates for our treatment dummies are consistently close to zero and corroborate the results from the non-parametric analysis—neither perceived urgency nor the variation of the post-purchase information structure affects average search length. In addition to these regression analyses at the search task level, we run Probit regressions for every stopping

[19] Calculating the number of safe choices in the risk elicitation task (Holt & Laury, 2002), participants are on average risk-averse. Meanwhile, 8.38 percent can be classified as risk-loving, 13.61 percent as risk-neutral. In the loss attitude task (Gächter, Johnson, & Herrmann, 2022), 4.71 percent of the participants maximize expected payoffs. While the fraction of participants accepting negative expected earnings is negligible (2.09 percent), the vast majority of the participants reject gambles with a positive expected value. The modal response is to accept gambles when the expected value of the gamble is larger than 2 EUR and reject them otherwise. Following the approach of Gächter, Johnson, and Herrmann (2022) we obtain a mean λ of 1.90 (with a standard deviation of 0.57), which is in line with recent literature (Brown et al., 2023). In the main regressions of Tables 1.3 and 1.4, we use a switching point to calculate the measures for risk and loss attitudes. Risk aversion is defined as the row when the participant switches from the safe to the risky lottery. Loss aversion is defined as the (inverse) row when the participant switches from accepting the risky lottery to rejecting it. For example, if a participant does not switch at all, then this is coded as 1. If a participant switches in row 1, then this is coded as 7. The results remain unaffected when we instead control for the number of safe choices (i.e., we take a measure that does not force the participant's responses to comply with monotonicity); see Appendix A.4.3.

decision within each search task (see Appendix Table A.4, Columns 1 and 2). This analysis confirms that treatments do not alter search length and shows in addition that decision-makers react systematically to prices. An increase in the current price by one unit approximately leads to a 1 percentage point decrease in the probability of accepting the current price offer. We thus provide robust and consistent evidence that treatments do not affect search length when considering all ten search tasks while, at the same time, decision makers take search costs systematically into account. We thus find no support for Hypotheses 2 but our evidence is in line with Hypotheses 5:

Result 2. *Considering all 10 search tasks, the number of requested offers does neither differ significantly between No-Info and Info nor between High-TP and Low-TP.*

1.5.4 Efficiency, Experiencing Regret, and Learning over Time

Next, we examine how efficient the search behavior is and how it evolves across the 10 search tasks. In total, 57.75 percent of the stopping decisions can be classified as optimal, in 26.60 percent of searches participants should have requested additional offers, and in 16.65 percent of the tasks participants searched too long compared to the reservation price of an expected payoff-maximizer. We observe minor differences across treatments. In *Low-TP*, 62.42 percent of the stopping decisions are optimal; in 24.11 percent of the tasks, too few offers are requested; and in 13.47 percent of the tasks, too many offers are requested. The fraction of optimal decisions in *High-TP* is lower than in *Low-TP* ($p = 0.001$, MWU) and amounts to 53.13 percent. In *High-TP*, the participants request too few offers in 27.08 percent of the tasks, and request too many offers in 19.79 percent of the tasks. Hence, behavior is slightly more diverse under *High-TP*. These differences translate into minor payoff differences (*High-TP*: 23.78 vs. *Low-TP*: 25.38; $p = 0.080$, MWU).

The fractions of optimal stopping decisions under *Info* and *No-Info* are closely aligned (*Info*: 57.37 percent vs. *No-Info*: 58.13 percent; $p = 0.879$, MWU) and payoffs do not differ substantially across the feedback conditions (*Info*: 24.43 vs. *No-Info*: 24.72; $p = 0.727$, MWU).[20] Under *No-Info*, in 24.84 percent of the tasks, more offers should have been requested; while in 17.79 percent of the tasks, fewer offers should have been requested. Similarly, in *Info* the fraction of tasks where too few offers were requested is 26.35 percent, and the fraction of tasks where too many offer were requested 15.52

[20]Because of this very similar efficiency across both feedback conditions, the lower efficiency under *High-TP* holds both without feedback ($p = 0.028$) and with feedback ($p = 0.016$).

percent. The closely aligned levels of efficiency across feedback conditions (*No-Info* and *Info*) may result from several reasons. First, participants may not consider the information provided and thus use similar decision processes in both information treatments. Second, participants may process feedback but not react (optimally) to it in subsequent tasks. Third, when participants are confronted with post-purchase information, they may change the overall sensitivity towards their own suboptimal behavior and react differently to similar information in *Info* as compared to *No-Info*. Concerning the first point, we avoided by design that participants simply ignored feedback, as in all treatments participants had to type in the (correct) number of the offer that would have yielded the highest payoff to proceed. Further, we do find evidence that participants spend substantially more time on the feedback screen in *Info* (25.53 seconds) as compared to *No-Info* (14.94 seconds; $p < 0.001$, MWU). It is thus unlikely that participants use similar decision making processes in both information treatments. To investigate the second and third point, we study experienced *inaction regret* (i.e. not having stopped early enough) separately in *Info* and *No-Info* and provide evidence on how experienced *action regret* (i.e., having stopped too early) alters search behavior in *Info* (where participants may learn that they have stopped too early).

Across all conditions, the participants experience *inaction regret* in 22.5 percent of the tasks. *Inaction regret* either arises due to the use of the recall option (79.59 percent of the cases in the data) or when the participants continue the search and encounter a better offer that still does not compensate for the additionally incurred search costs. While (experienced) *inaction regret* does not influence search behavior in general (see Table 1.4, Column 1), we find evidence that people in *Info* systematically react to the information provided as specified in Hypotheses 3 and 5 (see Table 1.4, Column 2). Knowing that one should have requested fewer offers in task k results in requesting around 1.14 offers less in task $k+1$ in *Info* compared to participants who did not experience *inaction regret*. In *No-Info*, experiencing *inaction regret*, if at all, slightly increases the number of requested offers (on average they request 0.47 offers more). That is, *inaction regret* (although possible in both treatments) affects subsequent behavior only in *Info*. This finding appears surprising but is consistent with an increased awareness towards regret feelings in general due to feedback provision in *Info*. In line with this idea, we observe that participants spend around 30% more time on the feedback screen in *Info* than in *No-Info* when experiencing *inaction regret* (*Info*: 22.37 seconds, *No-Info*: 17.17 seconds). Further, seeing additional prices in *Info* may reinforce inaction regret when the additional prices shown are inferior to the accepted price. We summarize this finding in Result 3:

Result 3. *Experiencing inaction regret in task k leads to a lower number of requested offers in task $k+1$ for participants in Info. For participants in No-Info, there is no such effect.*

Next, we assess how *action regret* influences subsequent search behavior. We first compare changes in search behavior in *Info* with changes in search behavior in *No-Info*. That is, we study search in task $k+1$, comparing participants in *Info* who requested too few offers from an ex-ante perspective and were informed by their feedback that they had stopped searching too early in task k with participants in *No-Info* who also requested inefficiently few offers from an ex-ante perspective in task k but did not see post-purchase prices that informed them about their inefficiently short search. For the regression analyses, we simulate the vector of prices participants in *No-Info* would have seen if they had been in the *Info* treatment (i.e., we randomly determine how many post-purchase price realizations they would have observed) and test for the effect of feedback on behavior in task $k+1$. At baseline (*No-Info/Low-TP*) in Table 1.4, Column 3, individuals average search length amounts to 5.02 offers. The average search length of individuals in *Info*, who experience *action regret* in t is increased by 1.1 offers. In contrast, participants in *No-Info* who also searched too short in task k and thus would have experienced *action regret* were they assigned to *Info* instead, continue to search too little (they request around 0.55 offers less in $k+1$). Column 4, which includes experienced *inaction* and *action regret*, and Column 5 which additionally includes interactions of both types of regret and time pressure, confirm these findings.[21]

As we randomly determined the number of displayed post-purchase prices within *Info*, we can also compare changes in behavior by participants within *Info* who requested too few offers from an ex-ante perspective and either were informed about having stopped too early and those who did not see more favorable post-purchase price realizations. We find that those who searched too short from an ex-ante perspective and

[21]The interaction between (experienced) *inaction regret* and time pressure in Column 5 implies that previous feelings of regret do not influence search behavior differentially when there is less time for deliberation. The constant coefficient for the interaction between *inaction regret* and *Info* also implies that the effect of time pressure after experiencing *inaction regret* is orthogonal to the *Info* treatment. The same holds true when adding the interaction between (experienced) *action regret* and *High-TP*. Here, the interaction term between *action regret* and *High-TP* suggests that participants who are under time pressure are somewhat more likely to search too short again (i.e., less likely to adjust their behavior). In Table A.2 in the Appendix, we show that the effect of experienced regret and time pressure is similar in both feedback structures.

Table 1.4: Experienced Regret

	Number of offers				
	(1)	(2)	(3)	(4)	(5)
Treatments					
High-TP	.203	.192	.211	.203	.432
	[-.295,.700]	[-.280,.664]	[-.265,.687]	[-.257,.663]	[-.103,.967]
Info	-.043	.189	-.335	-.106	-.115
	[-.497,.410]	[-.285,.664]	[-.828,.157]	[-.602,.389]	[-.610,.380]
High-TP X Info	-.172	-.129	-.200	-.155	-.135
	[-.821,.477]	[-.769,.511]	[-.847,.447]	[-.800,.491]	[-.771,.500]
(Experienced) Inaction Regret	-.082	.473		.420	.470
	[-.497,.332]	[-.127,1.073]		[-.161,1.002]	[-.228,1.168]
Inaction Regret X Info		-1.135***		-1.086***	-1.082***
		[-1.885,-.384]		[-1.816,-.356]	[-1.815,-.349]
Inaction Regret X High-TP					-.104
					[-.858,.650]
(Experienced) Action Regret			-.553*	-.513*	-.124
			[-1.111,.006]	[-1.062,.036]	[-.806,.558]
Action Regret X Info			1.095**	1.060**	1.058**
			[.239,1.951]	[.217,1.903]	[.225,1.891]
Action Regret X High-TP					-.757*
					[-1.598,.084]
# Tasks encountered	.065**	.068**	.061**	.065**	.064**
	[.009,.121]	[.012,.123]	[.006,.116]	[.010,.120]	[.008,.119]
Risk Aversion	-.066	-.065	-.069*	-.068*	-.065
	[-.145,.013]	[-.144,.014]	[-.147,.010]	[-.148,.012]	[-.146,.015]
Loss Aversion	.046	.049	.048	.053	.054
	[-.084,.176]	[-.079,.177]	[-.079,.175]	[-.075,.180]	[-.075,.183]
Constant	4.858***	4.666***	5.017***	4.821***	4.667***
	[3.505,6.210]	[3.297,6.034]	[3.601,6.432]	[3.380,6.262]	[3.229,6.106]
Socio-demographic controls	Yes	Yes	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes	Yes	Yes
Observations	1719	1719	1719	1719	1719

OLS Regressions.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. Values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching.Columns (1)-(4) display search behavior in tasks 2-10 and investigate the effect of regret experienced in the previous task. All columns include socio-demographic controls (gender, age, cognitive ability) and price sequence group fixed effects. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *Inaction Regret X High-TP* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *High-TP*. *Action Regret X High-TP* is defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 2-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

were informed about stopping too early requested on average 0.94 offers more in the subsequent task as compared to those who searched too short but did not see favorable post-purchase price realizations (3.69 vs. 2.75 offers requested in task $k+1$ after stopping too early in task k; $p = 0.058$, MWU). Result 4 summarizes these findings:

Result 4. *Experiencing action regret in task k leads to a higher number of requested offers in task $k+1$.*

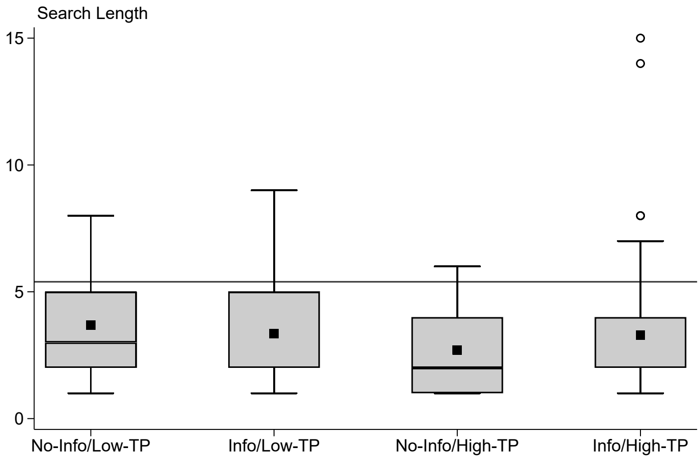
Although participants adjust their search behavior directionally, they do not make higher profits after experiencing regret in the previous task (see Appendix Table A.3, Column 1). This is true for both *inaction regret* and *action regret*. Participants who received information that higher earnings were possible had they stopped later (i.e., participants experiencing *action regret*) react by requesting inefficiently many offers in the next task. Table A.3 shows that the likelihood that participants continue to request too few offers remains unaffected (see Column 3), while the likelihood to ask for too many offers increases at the expense of optimal searches (see Columns 2 and 4).[22] Thus, we find evidence that participants react to experienced regret, but do not react optimally and, at the same time, that participants are more sensitive to information about *inaction regret* when experiencing the latter in *Info*.

Finally, we shed light on learning over time in terms of (sub)optimal choice. In the first half of their sequential search (tasks 1-5), the participants request on average around 1.57 fewer offers than ex-ante optimal ($p < 0.001$, Wilcoxon signed-ranks test). That is, suboptimal choice results mainly from stopping too early (participants request too many offers in only 15.39 percent of the first five tasks). Over time, participants request more offers (as shown by the *# Tasks encountered* coefficients in Tables 1.3 and 1.4) such that in the second half (tasks 6-10), the difference of the average search length to the optimal search length amounts to only 0.26 fewer offers than ex-ante optimal and does no longer significantly differ from the optimal benchmark ($p = 0.352$, Wilcoxon signed-ranks test). Overall, the fraction of searches where participants requested too few offers decreases from 36.13 percent in the first half to 15.08 percent in the second half, while the fraction of search tasks in which participants requested too many offers remains fairly constant (15.39 percent to 17.91 percent) across all treatments (see Figure A.2 in the appendix).

1.5.5 Anticipated Regret and Inexperienced Decision-makers

To isolate the effects of anticipated regret (excluding any experienced regret) and to study the effects of time pressure for inexperienced subjects, we now focus on the first task decision-makers encounter. Similar to our overall finding, participants stop also significantly earlier than optimal in the very first task (in all treatments, see Figure 1.4). While expected payoff maximizing behavior in the very first task results in stopping after seeing on average 5.39 offers, participants observe on average 3.26 of-

[22] In a robustness check (Table A.9), we show that all results hold in a truncated Poisson specification.



Notes. The figure shows boxplots of search lengths across treatments and a vertical line that indicates the optimal (ex-ante) threshold of a risk-neutral regret-free participant. The length of the whiskers is 1.5 times the interquartile range. The mean search length of each treatment is indicated by a solid square. The vertical line within the box corresponds to the median, which coincides with the lower quartile (lower end of the box) for Info/Low-TP and Info/High-TP.

Figure 1.4: Search Length across Treatments (Task 1).

fers. This difference is statistically significant when pooling the treatments and when analyzing them individually ($p < 0.001$ for each individual as well as the pooled test, Wilcoxon signed-ranks test). Search lengths in *Info* and *No-Info* are statistically indistinguishable ($p = 0.805$, MWU), while participants under time pressure search significantly shorter than participants without ($p = 0.019$, MWU). We corroborate the non-parametric analysis by regression analyses (see Table 1.3; Columns (4)-(6)). The results remain robust when adding demographic controls, and also when using independently elicited preferences as additional controls (Column 5) and when including price sequence group fixed effects (Column 6). Hence, also for the very first task we find no effects of the feedback environment.

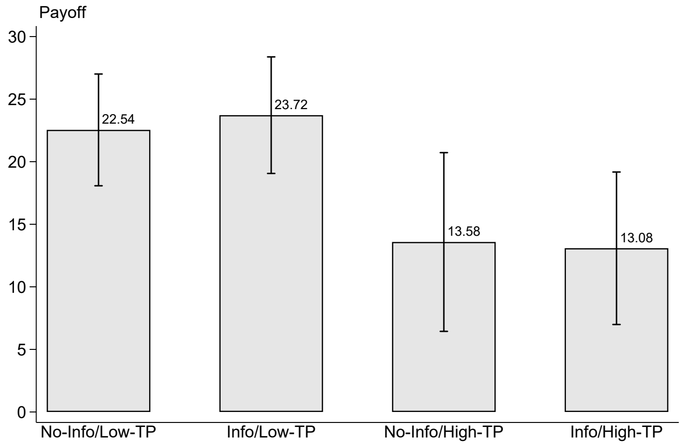
In contrast to our overall result, we do find a strong and statistically significant effect of time pressure on search length in the very first task (see also Table 1.3; Columns (4)-(6)), which substantially reduces payoffs in *High-TP*. As shown in Figure 1.5, under *Low-TP*, average payoffs amount to 23.14 Taler whereas in *High-TP*, participants' payoffs are more than 40 percent lower (on average they achieve only 13.33 Taler, $p = 0.004$, MWU).[23]

It is noteworthy, that perceived urgency was detrimental in the sense that subject in

[23] This comparison already excludes the extra cost that participants incurred in *High-TP* when exceeding the time threshold.

High-TP would not have fared worse when taking more time (as their counterparts in *Low-TP* did). When taking punishment costs due slower search in *High-TP* into account and applying the same punishment rule hypothetically to participants in *Low-TP*, our data suggests that, if at all, participants could have benefited from making slower choices. Hypothetical payoffs under *Low-TP* (with added costs for exceeding the threshold of 4 seconds) amount to 16.63 whereas those under under *High-TP* amount to 11.75 (when substracting the punishment costs for slow decisions; $p = 0.305$, MWU). Hence, ignoring the imposed time pressure and acting as if it was absent would have been at least as good in terms of payoffs as the strategies participants in *High-TP* resorted to.

Further, we provide additional evidence that participants reacted to pressure in a suboptimal way in the very first task, by comparing the number of requested offers conditional on the decision times in *Low-TP*. Note that the mere fact of deciding quickly does not imply short search durations in treatment *Low-TP*. Instead, swift decision-making is associated with a larger number of requested offers (Spearman's $rho = -.37$; $p < 0.001$). In efficiency terms, swift responses do not seem to be related to lower payoffs in *Low-TP*. Participants in *High-TP* who decided within 4 seconds perform substantially, although not significantly, worse (28.09 percent smaller payoffs; $p = 0.964$, MWU) than those who took more time to reach the decision (including the deduction for violating the time threshold). We interpret this as suggestive evidence that participants who (inefficiently) comply with the time threshold in the *High-TP* treatment by making faster choices than they would without time pressure do so in a systematic way (i.e., by requesting significantly fewer offers).

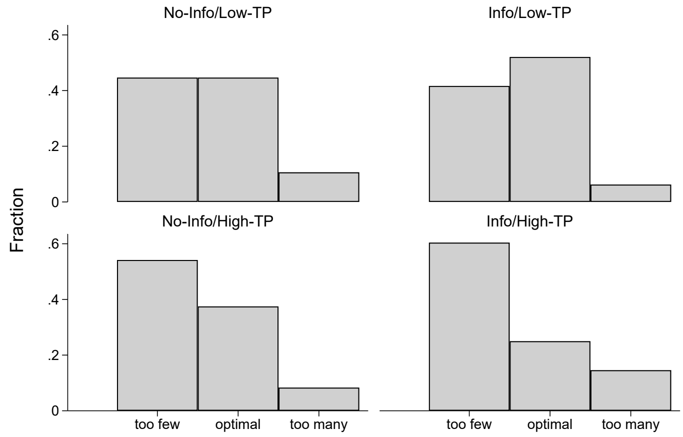


Notes. The figure shows the payoffs (in Taler) from the very first search task, excluding potential deductions for exceeding the time limit in High-TP conditions. The error bars indicate 95% confidence intervals.

Figure 1.5: Payoffs across Treatments (Task 1)

Summarizing the results for the very first task, we confirm the previously reported Results 1 and 2. Participants request significantly fewer offers in *No-Info* than the risk-neutral and regret-free benchmark predicts (Figure 1.6) and there is no (pure) effect of anticipated *action regret* on search behavior. In contrast to the analysis including all tasks, we find a significant effect of time pressure for the first task, which aggravates the existing tendency to request fewer offers than optimal. The latter is also confirmed in additional regression analyses considering every single stopping decision within the first task (see Table A.4, Columns 3 and 4 in the Appendix, which highlight that time pressure makes participants 15 percentage points more likely to stop the search at the current offer).

Result 5. *Participants request significantly fewer offers under High-TP than under Low-TP in the first search task they encounter, forgoing on average more than 40 percent of profits.*



Notes. The figure shows search behavior in Task 1. Behavior is classified as having requested either too few, too many, or the optimal number of offers compared to the (ex-ante) optimal behavior of a risk-neutral regret-free participant.

Figure 1.6: Ex-ante Efficiency of Search Behavior (Task 1)

1.6 Discussion

Our main findings documented limited differences across decision environments that may (or may not) involve anticipated regret and perceived urgency. In this section, we i) provide further insights into more subtle changes in search behaviors across decision environments and discuss how participants perceive their own decision quality (Section 1.6.1), ii) study whether commitment can serve as a simple tool to improve search efficiency (Section 1.6.2), and iii) provide evidence on the robustness of our result regarding the insensitivity to anticipated regret (Section 1.6.3).

1.6.1 Search Heuristics and Perceptions of Search Environments

Our main analyses focused on search behavior in the experiment compared to the risk neutral optimal benchmark (i.e., a constant reservation price). At the same time, previous literature (e.g., Hey, 1982; Moon & Martin, 1990; Houser & Winter, 2004; Schunk & Winter, 2009) highlights the importance of heuristics given the complexity to derive the optimal stopping rule in search tasks. Although our experimental design does not allow us to study all candidate heuristics discussed in this important previous work, we shed more light on individual search behavior related to i) salient stopping prices, ii) bounce-heuristics, and iii) streak-based heuristics across treatments (see Appendix Section A.6 for the details). We find that perceived urgency reduces the probability of acceptance of salient favorable prices ($p \leq 10$) and increases the probability of acceptance of salient unfavorable prices ($p > 50$) whereas our information conditions do not alter these probabilities. Further, we show that the use of bounce-heuristics does not strongly differ across information and time pressure conditions. For example, when analyzing the one-bounce heuristics following Houser and Winter (2004) and Schunk and Winter (2009) (i.e., "Have at least 2 searches and stop if a price quote larger than the previous quote is received."), we find that overall about 11 percent of decisions are consistent this heuristic, but this fraction does not strongly differ across treatments. Similarly, our additional analyses on streak-based heuristics (see Appendix Table A.14) results in minor treatment differences.

While overall treatment differences in search behavior appear minor, inexperienced participants do suffer from urgency. In turn, it is important to ask whether participants are aware of the influence of time pressure on decision quality. To study these perceptions in more detail, we elicit how decision-makers rank their performance as compared to other buyers at the end of the experiment. We find that on average, participants are overconfident in all treatments. Ranking themselves within a group of six (who all observed the same price sequences), they place themselves, on average, around one rank better than they actually are.[24] Although we do not find strong differences in actual performance across treatment when considering all 10 tasks, in a within-subjects comparison the participants expect to perform worse under time pressure than without ($p < 0.001$, Wilcoxon signed-ranks test). The difference is around 0.38 ranks on average. Although this holds for participants in both urgency conditions, it is stronger for participants who actually experienced time pressure in *High-TP* ($p = 0.018$, MWU test for differences in differences in rankings, comparing those assigned to *High-TP* and *Low-TP*, see also Figure A.1).

1.6.2 Improving Search Behavior through Commitment

Our study documents inefficient search across all treatment conditions and detrimental effects of time pressure for inexperienced decision-makers. Thereby our findings highlights the need for strategies consumers may employ to protect themselves. One simple strategy that may circumvent suboptimal search is commitment to a reservation price. In two additional search tasks, we explicitly asked participants to commit to a

[24] We do not neither observe significant differences in between treatments *No-Info* and *Info* ($p = 0.165$, MWU), nor between *High-TP* and *Low-TP* ($p = 0.959$, MWU).

reservation price instead of searching sequentially. We asked for such pre-commitment once with low (c=2) and once with high (c=4) search costs and compare their outcomes to their sequential search behavior.[25] Based on the reservation price stated for low and high search costs and realized prices in the sequential search tasks, we calculate when participants would have stopped the sequential search (if they had adhered to their stated reservation price). Doing so, we compare how the reservation price strategy fares with the same price sequence and with the same search costs as compared to sequentially requesting offers.

We find that commitment improves search efficiency. The percentage of optimal searches is significantly higher with pre-commitment than in the corresponding tasks of the main experiment (70.42 percent vs. 49.74 percent for search costs of c=2 and 80.10 percent vs. 67.02 percent with search costs of c=4; $p < 0.001$ for both search cost parameters, Wilcoxon signed-ranks tests). Hence, average reservation prices with pre-commitment are still above the rational benchmark, but the tendency to systematically request too few offers in early tasks is much less pronounced. Consequently, the participants achieve significantly larger profits with commitment (29.58 vs. 26.61 Taler for search costs of c=2, 21.49 vs. 20.09 Taler for search costs of c=4; $p < 0.001$ and $p = 0.014$, Wilcoxon signed-ranks tests).[26]

Note that this within-subject comparison does not allow us to rigorously disentangle effects of the different decision environment [choice of reservation price (pre-commitment) vs. sequential search] and learning over the experiment (because the tasks with pre-commitment followed after the 10 search tasks). However, we find that efficiency in the two tasks with pre-commitment is higher than in the last two of the 10 sequential tasks (13.48 percentage points more optimal decisions), suggesting that learning alone cannot explain the differences between the sequential search tasks and the tasks with pre-commitment.

To further disentangle learning and the effects of pre-commitment, we replicated the

[25] Reassuringly for our analyses of the value of pre-commitment, we find no indication that the treatments in the 10 sequential search tasks had an effect on search behavior in the additional search tasks with pre-commitment. This holds true when comparing the behavior in the two tasks separately ($p = 0.529$ for *High-TP* vs. *Low-TP* and $p = 0.883$ for *Info* and *No-Info* for Task 11 (c=2), MWU; $p = 0.914$ for *High-TP* vs. *Low-TP* and $p = 0.167$ for *Info* and *No-Info* for Task 12 (c=4), MWU) and jointly ($p = 0.61$ for *High-TP* vs. *Low-TP* and $p = 0.708$ for *Info* and *No-Info* for the average reservation price, MWU). In addition, we observe that participants reacted systematically to the incentives that they faced in the tasks with pre-commitment, choosing significantly higher reservation prices with high (as compared to low) search costs ($p < 0.001$, MWU).

[26] This remains unchanged if we only consider treatments without time pressure ($p < 0.001$ and $p = 0.007$ for search costs of c=2 and c=4, Wilcoxon signed-ranks tests).

two pre-commitment search tasks in an additional sample, in which participants did not encounter the ten sequential search tasks at all.[27] Again, we find support for the efficiency-enhancing effect of pre-task commitment. Reservation prices in the additional experiment that excluded learning possibilities do not differ significantly from reservation prices in the original experiment ($p = 0.405$ and $p = 0.923$ for search costs of c=2 and c=4, MWU). Moreover, we find that reservation price choices in the additional experiment lead to optimal stopping more often than sequential search behavior (with and without time pressure) in the 10 tasks of the main experiment (67.55 percent vs. 57.75 percent jointly; $p = 0.007$, MWU, with time pressure: mean = 53.13, $p = 0.002$, without time pressure: mean = 62.42, $p = 0.071$). Because the participants learned over time in the main experiment (as shown in Section 1.5.4), the difference is even more pronounced when comparing reservation price choices (which excluded learning possibilities) to the choices made in the first half of the main experiment (64.54 percent vs. 48.48 percent; $p = 0.001$, MWU).

1.6.3 Robustness: Non-binding Reservation Prices

Active sequential search takes place in different search environments. In many tasks, consumers are faced with the decision of whether to buy the product at a certain price or continue the search. In other tasks, the decision is characterized by setting a maximum acceptable price for the product, and continuing the search if the price was above this threshold (then with a potentially different new reservation value). Theoretically, both decisions are equivalent. Consumers *should* buy the product as soon the price is below their reservation value, irrespective of whether they first see the offer and then decide about buying or not, or whether they first specify their reservation value, and only then learn the value of the next offer. However, from a behavioral economics perspective, these decisions may be perceived differently. Asking repeatedly which future prices are acceptable (reservation value elicitation) may render behavior more future-oriented and thus alter the importance of anticipated regret. For example, recent evidence from Jhunjhunwala (2021) indicates that regret may play a more important role when repeated reservation value elicitation is used. In an additional pre-registered experiment (see AsPredicted; #80046), we investigate whether our findings are robust to such repeated reservation value elicitation. This experiment includes our treatments *No-Info* and *Info* without time pressure as baseline, as

[27] We recruited 47 subjects from the same pool as in the initial experiment (excluding all participants of the main experiment) and ran the additional sessions at MELESSA in September 2020.

well as two treatments in which participants set an initial reservation price before the first offer is drawn (*Reservation/Info* and *Reservation/No-Info*). If the first offer drawn is below the initial reservation price, the product is bought at the offered price. If the first offer is higher than the initial reservation value, participants can adjust their reservation value, and another offer is drawn for which search costs are incurred.

Interestingly, we find that information about post-purchase price realizations does not alter search behavior in an environment with non-binding reservation prices, either. Average search lengths are indistinguishable (*Reservation/Info*: 2.88 vs. *Reservation/No-Info*: 2.93; $p = 0.719$, MWU). Also the fraction of optimal stopping decisions under *Reservation/Info* and *Reservation/No-Info* is nearly identical (*Reservation/Info*: 61.88 percent vs. *Reservation/No-Info*: 60.00 percent), translating into very similar payoffs (*Reservation/Info*: 23.27 vs. *Reservation/No-Info*: 23.48; $p = 0.877$, MWU). In Appendix A.5, we discuss these findings in more detail and compare search behavior under both elicitation procedures. Further, we show that the exact replication of the baseline treatments (*Info* and *No-Info* without time pressure) confirms the negligible role of anticipated post-purchase information on search length and payoffs observed in the main experiment.

1.7 Conclusion

From a theoretical perspective, perceived urgency and regret may substantially affect individual choice in dynamic market environments and hence aggravate or alleviate any potential biases in decision-making. We used a well-powered experimental study to evaluate the empirical importance of both aspects and their interaction. The 95 percent confidence intervals for the treatment effect estimates in our preferred regression specification (Table 1.3, Column 3) are consistent with differences across treatments of up to 0.66 requested offers, corresponding to 0.17 standard deviations in the number of requested offers. Hence, we can rule out true but undetected effect sizes being larger than 0.17 standard deviations. We obtain very similar results when deriving minimum detectable effect sizes using a simulation-based approach (see Campos-Mercade, 2018). Based on the realized distribution of search lengths, we set the desired level of power to 80 percent and the statistical significance level to 5 percent. We then perform parametric and non-parametric tests, and find that we are able to detect effect sizes of at least 0.15 standard deviations across all tasks. Hence, our study ex-ante allowed us to detect economically meaningful treatment differences and thereby allows is to study the role of time-pressure, regret, and their potential interaction in sequential search tasks.

First, we find that urgency significantly affects search behavior and profits in the very first search task that the participants encounter. Under high time pressure, stopping too early is (even) more prevalent than under low time pressure and profits are substantially reduced. Thus, our results provide one rationale for why sellers often put buyers under time pressure. Clearly, short-lived discounts can deter search, since they limit consumers' ability to consider alternative offers before the discount expires. Search deterrence can be even more pronounced if sellers can discriminate against buyers who do not purchase at the first opportunity (Armstrong & Zhou, 2016). Our findings additionally emphasize a channel of bounded rationality. Pressuring buyers by inducing a sense of urgency may be particularly effective when applied to inexperienced customers (i.e., customers who have not encountered the respective search task before). With experience, participants in our experiment were not significantly affected by time pressure. Consumer protection policies against sales tactics that "rush consumers into making a decision",[28] can thus be especially helpful for inexperienced consumers. These are, for example, customers who are in an environment where they are not very savvy, or who are searching for products that they have not previously looked for. For example, the British Competition and Markets Authority recently required booking sites to take action against practices of pressure selling (i.e., practices that create perceived urgency) and of displaying potentially misleading unattainable offers [i.e., that give rise to (anticipated) feelings of regret], such as already forgone options. Given that booking flights or hotels is a regular task for many consumers, they may quickly learn to resist the sense of urgency and make better decisions. However, other purchase decisions may be more infrequent but substantially more important. Buying a house, taking out life insurance, or making other long-term investment decisions presents most consumers with an unknown decision environment. As we find that perceived urgency particularly harms decision quality of inexperienced participants, regulation may be more important in such 'unknown' environments than in areas that are currently primarily targeted (e.g., hotel booking or travel websites).

Second, our results provide robust empirical evidence that anticipated regret does not generally affect the number of requested offers in sequential search tasks. In particular, we do not find that anticipated regret renders active sequential search. While avoid-

[28]Retrieved from https://www.gov.uk/government/news/cma-launches-enforcement-action-against-hotel-booking-sites on 10/05/2020.

ing anticipated regret (Bell, 1982; Loomes & Sugden, 1982; Skiadas, 1997; Hayashi, 2008; Sarver, 2008; Bikhchandani & Segal, 2014; Qin, 2015; Halpern & Leung, 2016; Buturak & Evren, 2017) has been observed in other experimental contexts (Zeelenberg, 1999; Camille et al., 2004; Coricelli et al., 2005; Strack & Viefers, 2021; Fioretti, Vostroknutov, & Coricelli, 2022), such regret seems to play a minor role when decision-makers incur salient search costs by actively requesting new price offers and learn advantageous and disadvantageous post-purchase prices. We replicate this result in an additional experiment (see Section 1.6.3) and show that the observed insensitivity towards anticipated regret in our setting does not hinge on whether participants directly chose to buy the product or repeatedly specify reservation prices. Recent evidence by Jhunjhunwala (2021) suggests that such search behavior can be affected when the feedback structure only highlights potentially better offers. In contrast, in our setting participants see a random subset of actual future price realizations (and associated payoffs). Thus, our results show the tight boundaries of changes in search behavior through post-purchase information: in a search environment where consumers may learn (a subset of) all competitors' prices post purchase, changes in search behavior due to anticipated regret appear unlikely while in environments, where consumers anticipate to only see prices that provide a better deals (e.g., because competitors may be more likely to advertise such prices), anticipated regret may result in searching longer. There are two reasons that may explain why we do not identify strong efficiency effects of regret. First, anticipated *action regret* might not have been very salient for participants because the recall option makes the subjects perceive that good deals are still available, although net benefits from trade are much smaller when searching longer due to search costs. Furthermore, explicit search costs, as well as the fact that a new price requires an active choice, may render the search-prolonging role of anticipated regret less salient. Second, regret might have been salient but the decision environment was too complex to allow for efficiency-enhancing effects. Our results are in line with a combination of both explanations. In the very first task that the participants encounter, anticipated regret plays a minor role (in line with anticipated regret not being very salient); whereas participants who received post-purchase information still reacted to experienced *action* and *inaction regret*. However, participants were not successful in making better decisions in subsequent search tasks with different price realizations and search costs.

Third, we do not find a substantial interaction between anticipated regret and perceived urgency. Independent of the decision environment, our results indicate that individuals search too little. However, our results also hint at a simple mechanism that consumers may use to avoid such inefficient search: commitment to a binding reservation price. In the experiment, commitment increases search length and payoffs. As such, commitment may also be applied as a potential solution outside the laboratory, and sophisticated consumers may demand commitment devices in the form of public policies or market-based solutions.

Finally, although our design captures many important elements of the trade-off that urgency, resulting time pressure, and regret in real-world settings may pose, decision environments outside the laboratory may both confront consumers with additional challenges or relieve them of some that exist in our setting. On the one hand, repeated search tasks in which search costs stay constant may render learning from past experiences and regret easier (see e.g., Einav, 2005; Oprea, Friedman, & Anderson, 2009) while more complex environments with varying search costs (as in our experiment) may render learning harder. On the other hand, in many search environments outside the laboratory, consumers face uncertainty about the underlying distribution from which prices are drawn and firms may have an incentive to disguise certain pieces of information to create more intransparent decision environments, thereby complicating optimal search. Hence, understanding in greater detail how the aversive feelings of regret and urgency connect to actual decision quality in different environments seems a promising route for future research.

2

Everyone Likes to Be Liked

Experimental Evidence from Matching Markets

2.1 Introduction

We often prefer to interact with individuals who also want to interact with us. For example, applicants may reconsider a job offer after learning they were not the first-choice candidate (Antler, 2019).[1] They may realize that an employer who does not favor them will be less invested in their relationship. Or they may be less willing to invest in such a relationship themselves. We say that individuals who prefer to be matched with a partner who wants to be matched with them have *reciprocal preferences*.

Reciprocal preferences seem to play a crucial role in matching markets, where partici-

*This chapter is based on joint work with Christoph Schwaiger.

[1] See also https://www.forbes.com/sites/lizryan/2018/01/20/im-the-second-choice-candidate-should-i-still-take-the-job, accessed 07/18/2022.

pants cannot unilaterally choose their partner, but must also be chosen. For example, schools seek to know students' preferences to take into account "who wants them most" when making admission decisions.[2] When the preferences of other market participants are not disclosed, agents even modify mechanism to attract those who want to be matched with them. To achieve this, it became common practice for German universities to only accept medical students who had ranked the respective university favorably.[3] Similarly, Avery and Levin (2010) show that universities use early admission programs to admit highly interested students who, in turn, have lower grade point averages. Such policies, while individually rational, undermine the efficient functioning of matching markets. Opitz and Schwaiger (2023b) show theoretically that reciprocal preferences even cause agents to break up their assigned match when centralized matching mechanisms are in place –contradicting the main objective of matching mechanisms to establish stable relations.

In this study, we identify reciprocal preferences and their impact on matching markets through a laboratory experiment. In the experimental setting, we observe participants' preferences under different information sets. In contrast to observational data where neither the true preferences of market participants nor their information sets are known precisely, this allows us to identify reciprocal preferences. We directly test whether agents' preferences are sensitive to information about others' preferences. We hypothesize that participants *like to be liked*. That is, they prefer a partner who ranks them favorably (Aronson & Worchel, 1966; R. M. Montoya & Horton, 2012). Therefore, participants change their preference order after learning how others ranked them, which leads to instability in the matching market.

In the experiment, participants form two-person teams for a Public Goods Game (PGG) through a centralized matching mechanism. During the team-formation stage, participants interact in groups of eight, evenly divided between two sides of the market. Based on personality questionnaire responses, participants indicate with whom from the other market side they would like to play the PGG. They submit a rank-ordered list of potential partners from the other market side to a centralized Deferred Acceptance (DA) mechanism. The DA mechanism theoretically achieves stable allocations in two-sided matching markets, such that no participant benefits from breaking up

[2] Concerns were raised about changes in the school admission system that left principals uninformed about students' rankings of the schools (see https://www.nytimes.com/2004/11/19/education/council-members-see-flaws-in-schooladmissions-plan.html, accessed 07/18/2022).

[3] While this practice was prohibited by the Federal Constitutional Court in 2017 (*BVerfG, 1 BvL 3/14*, 2017), many institutions still have similar procedures, such as Trinity College in Toronto, which only accepts students who rank the college first.

the formed match (Gale & Shapley, 1962). In our treatment (*Info*), one side of the market receives information about with whom they are tentatively matched and how their potential partners rank them. In our baseline (*No-Info*), this market side never learns how their potential partners rank them and only sees with whom they are tentatively matched. In both treatments, they can subsequently change their preference list, resubmit it to the mechanism, and may get a new partner as a result. Afterwards, the matched partners play a standard PGG designed to capture the essential trade-off between collectively beneficial but individually costly contributions to a public good. This design allows us to understand the effects of reciprocal preferences on the stability of matching markets.

We develop a stylized model to study two possible channels for the emergence of reciprocal preferences in cooperative settings. The first channel is belief-based. It assumes that agents expect partners who like to be matched with them to be more cooperative (i.e., they expect their partner to contribute more in the PGG). The belief that favorable preferences signal a higher match-specific payoff provides a monetary rationale for reciprocal preferences. The second channel is preference-based. This channel posits that agents derive higher utility from the material well-being of a matched partner who likes them. As a consequence, they prefer to be more cooperative themselves (i.e., they contribute more in the PGG). Both channels imply that being matched with a partner who ranks the agent favorably spurs a higher utility, thereby providing a foundation for reciprocal preferences. The experiment allows us to test both channels.

Our outcome variables stem from the team-formation stage and the PGG. The first set of outcome variables investigates the effect of reciprocal preferences on stability. Achieving stable outcomes is central to matching mechanisms and implies Pareto efficiency (Gale & Shapley, 1962); Opitz and Schwaiger (2023b) show that reciprocal preferences can lead to instability when agents update their beliefs about others' preferences after the allocation of the mechanism.[4] We analyze whether participants change their preference order once they learn how they are ranked by their potential partners, whether these preference changes are indicative of reciprocal preferences, and how this affects stability. The second set of outcome variables is based on subsequent behavior in the PGG and sheds light on belief-based and preference-based

[4]Updating can either happen through directly learning others' preferences (as in the experimental design), or more subtly through observing the final matching and being able to make inferences about the underlying preferences that led to the matching. For a detailed analysis, see Opitz and Schwaiger (2023b).

channels underlying reciprocal preferences. We test whether reciprocal preferences are belief-based by eliciting incentivized beliefs about the partner's contributions to the PGG. To test the preference-based channel, we focus on conditional contribution decisions. In these decisions, we isolate altruism from the beliefs about a partner's contribution.

The main results are as follows: First, agents adjust their preferences significantly more often when they observe their potential partners' preferences (*Info*) than when they do not (27.67 vs. 9.67 percent). These preference adjustments in *Info* are consistent with reciprocal preferences. Participants rank those who rank them favorably higher than those who do not - they *like to be liked*. Second, these preference adjustments translate into significantly more unstable matchings in *Info* than in *No-Info* (40.00 vs. 10.67 percent). This provides strong evidence that reciprocal preferences can inhibit the desired functioning of matching mechanisms. Third, our results indicate that both belief-based and preference-based motivations underlie reciprocal preferences. We show that participants hold (accurate) beliefs that someone who likes to be matched with them will be more cooperative. In this sense, revealed preferences signal the future value of the match, providing a profit-maximizing rationale for working with someone who likes you. In addition, we find evidence that participants act more altruistically towards those who indicated a preference towards them, providing support for a preference-based foundation. Lastly, in *Info*, we document higher average cooperation and profits.

Our findings contribute to a better understanding of matching markets, team formation, and team behavior. First, we contribute to the growing experimental literature on matching markets (see Hakimov & Kübler, 2021, for a review). This literature attempts to uncover factors that limit the efficient functioning of matching markets because they affect agents' strategies. We are the first to study reciprocal preferences experimentally, and investigate whether outcomes of the DA mechanism remain stable. In this way, we test the empirical stability of the DA mechanism when others' preferences are revealed. Closest to this is previous work on the impacts of information about other participants' preference profiles and reporting strategies in one-sided (Pais & Pintér, 2008) and two-sided centralized markets (Pais, Pintér, & Veszteg, 2011;

Shimada, 2022) on truth-telling.[5] These papers center on the extent to which agents use additional information to misrepresent their preferences strategically across mechanisms. In contrast, we are interested in the causal effect of knowing one's rank in the preference order of potential partners, and demonstrate how this reduces stability.

Second, we contribute to social preferences, team formation, and cooperation literature. Individuals often prefer to interact and team up with agents who are similar to them, which is known as *homophily* (McPherson, Smith-Lovin, & Cook, 2001). Homophily can be observed in experimental settings (Currarini & Mengel, 2016; R. Chen & Gong, 2018), as well as economic settings (e.g., the choice of co-workers and entrepreneurial teams (Hedegaard & Tyran, 2018; Boss et al., 2023)). Self-selected teams display higher satisfaction, collaborative spirit, and effort (R. Chen & Gong, 2018; Boss et al., 2023), while results on performance are mixed.[6] We contribute to the organizational literature on efficient team formation by highlighting the role of reciprocal preferences. We show that for an individual not only the similarity with their potential team partners matters, but also their partners' preferences.

Moreover, individuals are more likely to cooperate with those they perceive as similar. People are more cooperative if they perceive others to belong to the same group (Akerlof & Kranton, 2000), where social identity may either be fostered through previous interaction (Eckel & Grossman, 2005), or by a shared preference (e.g., Y. Chen & Li, 2009, with a minimal group paradigm). Consequently, social proximity can overcome market imperfections (Chandrasekhar, Kinnan, & Larreguy, 2018; Jain, 2020), leading to higher levels of altruism (Leider et al., 2009; Goeree et al., 2010). Given that similar people also like each other, our paper provides an explanation for why similarity leads to higher cooperation. This is in line with recent literature showing that mutual dislike often hinders team performance (Gerhards & Kosfeld, 2020).

Lastly, people also treat those who have been generous towards them more favorably (Akerlof, 1982). We extend the recent literature on reciprocity towards non-monetary gifts (Kube, Maréchal, & Puppe, 2012; Bradler et al., 2016). In our experiment, receiving a favorable rank can be interpreted as a non-monetary gift, which leads to higher cooperation. With this, we show that interpersonal preferences are another currency

[5] In Shimada (2022), experimental participants are matched with computerized players. When computerized players apply a strategy in which a participant moves up in their preference list if the participant evaluates them favorably (i.e., in our terminology they are programmed to have reciprocal preferences), participants misrepresent their preferences as a response. This is similar to theoretical results in Opitz and Schwaiger (2023b).

[6] See Horwitz and Horwitz (2007) for a broader discussion on homophily and (workplace) performance.

of reciprocity, most closely related to the idea in R. Dur (2009) that "employees care more for their manager when [...] their manager cares for them".

The paper is structured as follows: Section 2.2 presents our experimental design, Section 2.3 outlines our hypotheses and results on reciprocal preferences at the matching stage. Section 2.4 illustrates the underlying channels through a stylized model, and investigates these channels empirically. Finally, we discuss and conclude in Section 2.5.

2.2 Experimental Design

Research Questions Through our experimental design, we examine three main research questions. First, do participants have reciprocal preferences? Second, do reciprocal preferences lead to instability in matching markets? Third, what are the mechanisms underlying the change in stated preferences? To address these questions causally, we exogenously manipulate information structures between treatments. This provides us with the necessary variation that observational data cannot give us to identify reciprocal preferences, their underlying mechanisms, and their implications for matching markets.

Overview The pre-registered experiment consists of three main parts.[7] In Part I, we collect self-reported personality data. In Part II, participants form two-player teams through a centralized matching mechanism and play a PGG within the formed dyads. Participants indicate with whom they would like to team up based on their potential partner's personality profiles from Part I. In Part II, we compare behavior under two information structures in a between-subject design. In the treatment condition (*Info*), participants on one side of the market learn how their potential partners ranked them before submitting their final preference ranking. In *No-Info*, participants never know how their potential partners ranked them. In Part III, we elicit beliefs about the PGG contribution of their team partner and collect control variables (loss aversion/ cognitive ability/ gender). The design is visualized in Figure 2.1.

Part I All participants fill out a personality questionnaire with 15 items on a four-point Likert scale. It contains five statements each on personality traits, preferred leisure activities, and societal opinions (see Appendix B.5.2 for the complete ques-

[7] The preregistration of our design, as well as a detailed pre-analysis plan can be found at AEARCTR-0007551.

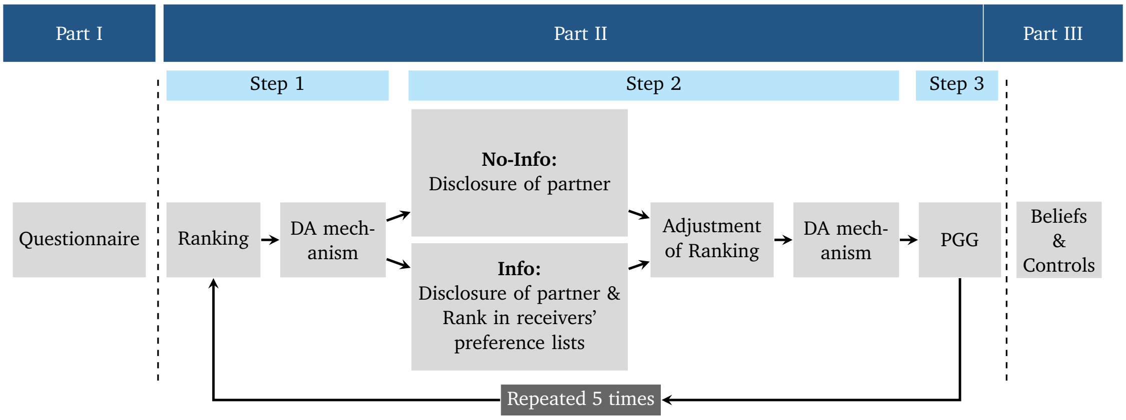


Figure 2.1: Design Overview

tionnaire). Since the later ranking is based on answers to these questions, they are intended to give an impression about the respondent.[8]

Part II Participants are randomly assigned to one of two market sides. As it is standard in two-sided markets, these roles are referred to as proposers and receivers.[9] In each *matching market*, there are four proposers and four receivers. The centralized DA mechanism (Gale & Shapley, 1962) forms four teams, consisting of one proposer and one receiver. We rely on the DA mechanism because of its theoretically desirable properties, in particular because the final allocation is stable under standard assumptions. This procedure is the same in every *matching market*. Part II consists of three steps.

Step 1 Proposers and receivers submit a rank-ordered list of their potential partners. Proposers rank the four receivers in their *matching market* according to the desirability to be matched with them and vice versa. Teams are tentatively formed through the centralized DA mechanism, which matches one proposer with one receiver for the upcoming PGG.[10] Participants submit their preferences based on questionnaire responses from Part I. Each proposer in the *matching market* sees the same five randomly chosen answers from each receiver. The receivers see the answers to five different questions

[8] At the same time, the answers should not provide clear information about their cooperation behavior to minimize the initial correlation of preferences on each market side. In the extreme case, every participant on one side of the market submits the same preferences to the mechanism. Then, *reciprocal preferences* do not affect the outcome, because all potential partners are equally inclined to cooperate with one.

[9] In the beginning of Part II, participants are informed about their role, and receive detailed instructions on the procedures of the team-formation process and the PGG (see Appendix B.5).

[10] This means that we study a setting of two-sided matching in a one-to-one matching market, often referred to as a *marriage market* following Gale and Shapley (1962).

randomly selected among the remaining ones.[11] After the participants submit their rank-ordered lists, the DA mechanism forms the tentative allocation.

Step 2 Proposers can submit a revised preference list to the DA mechanism. We vary the information between our two treatments *Info* and *No-Info*. In *No-Info*, proposers see with whom they have been matched in the first step. In *Info*, a proposer receives additional information on how all receivers ranked him. After examining this information, proposers decide on whether to revise their preference list and re-submit it to the DA mechanism. Receivers do not play an active role in this step as their preferences remain fixed. Furthermore, they never learn that proposers can adjust their preferences. Proposers know that receivers never learn about proposers' preferences (and changes thereof). The DA again forms a tentative allocation. Then, one of the two tentative allocations (Step 1 or Step 2) is implemented with equal probability.

Step 3 The formed dyads play a two-player PGG in the final step of Part II. Both partners receive an initial endowment of 10 Taler (experimental currency) to be either allocated to a private account or to be contributed to a public account. The contributed amount of each partner $c_i \in \{0, 1, ..., 10\}$ is referred to as the unconditional contribution. The sum of both players' contributions to the public good is multiplied by 1.5, and divided equally between the two. This leads to the following payoff function for a participant i: $\pi_i = 10 - c_i + 0.75 * (c_i + c_j)$. The marginal per capita return of 0.75 implies that free-riding ($c_i = 0$) is the dominant strategy from an individual perspective. However, since the sum of marginal returns is greater than 1, contributing the entire endowment of 10 Taler maximizes the team surplus. In addition to the unconditional contribution, proposers also fill in a table indicating their contribution for every possible contribution of their matched partner, referred to as their conditional contributions (Fischbacher, Gächter, & Fehr, 2001). Receivers only state their unconditional contribution.[12] The final payoff for the receiver depends on the stated unconditional contributions of both players. The final payoff for the proposer depends on the receiver's unconditional payoff and on the proposers conditional or

[11] The intuition for sharing distinct questions is to minimize the initial correlations between preferences across market sides. If similarity is a relevant determinant for the choice of a partner (*homophily*), different questions provide different information about similarity, which reduces the correlation of preferences. In the extreme of perfect correlation, everyone is already matched with the partner they prefer most and that prefers them most, such that *reciprocal preferences* do not affect the outcome.

[12] This circumvents the problem with conditional contributions that the standard (unique) Nash-Equilibrium of not contributing anything requires common knowledge of rationality (Fischbacher, Gächter, & Fehr, 2001, Footnote 6). In light of a substantial fraction of conditional cooperators in previous PGG experiments, we do not want to assume this and let receivers only make an unconditional contribution decision (which is known to the proposers).

unconditional contribution.

Part III We complement the contributions to the PGG with incentivized point beliefs about partner's unconditional contribution, for both proposers and receivers (Gächter, Kölle, & Quercia, 2017). We do not announce the belief elicitation before, to rule out that expectations about the ability to judge the behavior of another player influence preference submission.

We also elicit proxies for cognitive ability, loss attitudes, and socio-demographic controls. Proposers with higher cognitive abilities may be more likely to perceive receivers' preferences as signals for their contribution and adjust their preferences strategically. We use Raven's Matrices as a proxy for cognitive ability.[13] Participants are given 5 minutes to complete increasingly difficult Raven's Matrices, scored on the number of correct answers minus the number of incorrect answers. High degrees of loss aversion may make participants less likely to adjust their preferences if they feel attached to their current partner. Although unlikely given the information sets of participants in our experiment, (expectation-based) loss aversion may influence initial reporting strategies (Meisner & von Wangenheim, 2023). Hence, we elicit an incentivized measure of loss aversion in risky choices (Gächter, Johnson, & Herrmann, 2022). Before concluding the experiment, participants complete a short socio-demographic questionnaire.

Repetitions We repeat Part II five times. During each repetition, participants play within a new *matching market* of randomly selected participants. Roles as proposer or receiver remain constant across rounds. To minimize the influence of earlier rounds on later rounds, participants do not receive feedback between rounds. Furthermore, by displaying only a subset of questionnaire responses in each round and randomly assigning participants to *matching markets*, we minimize the possibility that participants may identify others across rounds.

Payoffs and Incentive Compatibility One round of the PGG is randomly chosen to be payoff relevant. Participants earn money through their final payoff from the PGG (determined by their own and their partner's contribution choice) in one of the five rounds. We randomize whether the conditional or the unconditional contribution decision of a proposer is implemented. Through the compensation in the PGG,

[13] The Raven's Matrices test is a leading non-verbal measure of analytic intelligence, test scores are associated with the degree of sophistication in the beauty contest (Gill & Prowse, 2016), in manipulable matching mechanisms (Basteck & Mantovani, 2018), as well as with more accurate beliefs (Burks et al., 2009).

we incentivize the submission of truthful rank-ordered lists. To guarantee that both the initial submission, as well as the potentially revised preference order are incentive compatible, one of the two is implemented with equal probability to determine the final matching. We incentivize the point beliefs about their partner's contributions. Participants receive a fixed amount if their stated belief corresponds to the actual unconditional contribution, and they receive no payment otherwise. Additionally, participants are paid based on their performance in the Raven's matrices task and the loss attitudes elicitation.

Experimental Procedures The experiment was conducted at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA). In total, 235 student participants participated in the experiment. The participants were recruited using the online system ORSEE (Greiner, 2015). The experiment was programmed with the software oTree (D. Chen, Schonger, & Wickens, 2016). We conducted 10 sessions (5 sessions per treatment, each with a desired number of 24 participants). On average, participants earned 21.5 EUR (including a show-up fee of 6 EUR). The experiment lasted about 80 minutes.

2.3 Reciprocal Preferences in Matching Markets

Our experimental design test the hypothesis that proposers adjust their preferences in *Info* to be matched with a receiver who wants to be matched with them, which in turn leads to a different matching outcome (instability). In pre-registered analyses, we test whether proposers adjust their preferences more often in *Info*, whether these adjustments lead to higher instability, and whether they display reciprocal preferences. Exploratory analyses that were not pre-registered are marked as such.

2.3.1 Instability of the Deferred Acceptance Mechanism

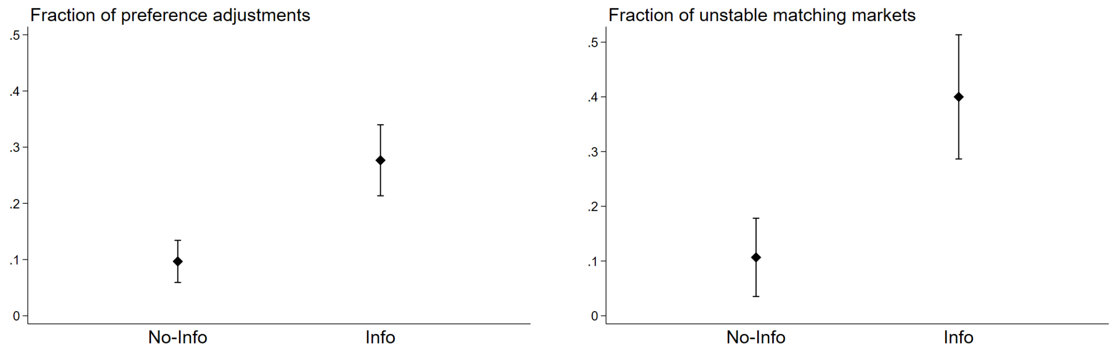
Proposers change their individual preferences more often when they see their potential partners' preferences (*Info*) compared to when they do not see the receivers' preferences (*No-Info*).

Result 1. *The fraction of preference adjustments in Info is 27.67 percent, while it is only 9.67 percent in No-Info. This difference is significant (*$p < 0.01$*; Mann-Whitney-U test (MWU)).*

Regression analysis in Table B.1 in Appendix B.1 confirms that the fraction of preference adjustments is significantly higher in *Info* (see also Figure 2.2, Panel A).

As a consequence of more frequent preference adjustments, the fraction of *matching markets* where the rematching outcome changes after the rematching stage is larger under *Info* (Figure 2.2, Panel B). Instability of a matching is defined at the *matching market* level. We compute the resulting matching with both the initial and the (potentially) revised preference list. A matching is stable when both resulting matchings are the same (i.e., if all participants are matched to the same partner). Otherwise, a matching is unstable. This implies that a matching market is unstable if at least one of the proposers changed their preferences list, and this change led to a different market outcome. A change in reported preferences leads to a different outcome only if it results in a proposer-receiver pair that prefers to be matched to each other compared to their current match. The fraction of unstable matchings is substantially larger under *Info* than under *No-Info*. Hence, a *matching market* is nearly four times more often unstable under *Info* than under *No-Info*.

Result 2. *There is significantly higher instability in Info than in No-Info. The fraction of unstable matching markets in Info is 40.00 percent; it is only 10.67 percent in No-Info ($p < 0.01$; χ^2 test).*



A) Preference Adjustments B) Instability

Notes. This figure displays the fraction of proposers who adjust their preference lists in Panel A, and the fraction of unstable matchings Panel B across the two treatments *No-Info* and *Info*. The vertical lines indicate the 95% confidence interval.

Figure 2.2: Reciprocal Preferences and Stability

Thus, we conclude that proposers are more likely to adjust their preference ranking when they see the preferences of receivers, leading to instability in the DA mechanism.

2.3.2 Reciprocal Preferences and Preference Adjustments

Proposer's preference changes indicate the presence of reciprocal preferences. For each preference adjustment, we can classify whether it is consistent with the participants having reciprocal preferences or not. A proposer's preference adjustment is *consistent* with reciprocal preferences if the now more favorably ranked receiver(s) gives a strictly better rank to the proposer compared to the now less favorably ranked receiver(s). Formally, this requires that if $Proposer\, P$ switches the position of $Receiver\, R$ and $Receiver\, S$, and $Receiver\, R$ was the initially more preferred candidate, then $Proposer\, P$ must have been ranked strictly better by $Receiver\, S$ than by $Receiver\, R$.[14]

Our results strongly support that preference adjustments largely reflect *reciprocal preferences*. In *Info*, 73.68 percent of the adjustments are consistent with *reciprocal preferences*. This compares to a fraction of 20.69 in *No-Info* where participants could not systematically react to others' preferences.[15] The difference between both conditions is significant ($p < 0.01$; MWU). Table B.2 in Appendix B.1 confirms these findings through a logit regression, documenting a significantly higher likelihood of a consistent preference adjustment (compared to an inconsistent adjustment or none) in *Info*, both in a uni-variate regression (Column 1) and when adding individual-level controls (Column 2).[16]

A more detailed exploratory analysis of the determinants of preference adjustments supports the conjecture that proposers' preference adjustments reflect *reciprocal pref-*

[14]For a formal introduction of reciprocal preferences into matching markets, we refer the interested reader to Opitz and Schwaiger (2023b). In Appendix B.4, we provide some theoretical intuitions for why preference changes consistent with reciprocal preferences indeed lead to instability.

[15]If participants switched the position of two receivers in the preference lists randomly, we would expect 20.9% of the adjustments to be consistent with reciprocal preferences by chance. 24 out of 29 preference adjustments in *No-Info* are such that (only) two receivers switch their position. In more complex cases, the probability of a random adjustment being consistent with reciprocal preferences is even lower.

[16]In the loss attitude task (Gächter, Johnson, & Herrmann, 2022) individuals are asked to choose between no payment and a risky lottery with one negative and one positive outcome. Every individual makes several decisions. We keep the positive outcome fixed at 6 Euro, the negative outcomes varies between a loss of 2 and 7 Euro. 2.55 percent of the participants maximize expected payoffs. While the fraction of participants accepting negative expected earnings is negligible (1.28 percent), the vast majority of the participants reject gambles with a positive expected value. The modal response is to accept gambles when the expected value is larger than 2 EUR and reject them otherwise. Loss aversion is defined as the lottery where a participant switches from accepting to rejecting it. For example, if a participant accepts all lotteries, this is coded as 1. If a participant accepts no lottery, this is coded as 7. Cognitive ability is calculated by the number of correctly solved matrices, minus the number of incorrectly solved ones. Out of 10 matrices, participants achieve an average net score of 6.23. 2.55 percent of participants did not solve any matrix correctly, while 5.53 percent solved all 10 matrices correctly.

erences (see Table B.5 in Appendix B.3). First, the more favorably a proposer ranks their initial partner, the lower the likelihood that the proposer will adjust preferences. This holds true both in *No-Info* (Column 1) and *Info* (Columns 2 & 3). Second, receivers' preferences matter when deciding whether to adjust the preference ranking in *Info*. Being liked by the (tentatively) matched receiver lowers the likelihood that a proposer adjusts their preferences. At the same time, being a preferred candidate by other (non-matched) receivers increases the likelihood of adjusting preferences. Column 2 shows that a more favorable average rank by the non-matched receivers increases the likelihood of adjusting the preference ranking; Column 3 confirms this pattern by estimating the effect of the best rank received by one of the other three receivers. That proposers in *Info* are less likely to adjust preferences when their matched partner ranked them favorably, and more likely when the other potential partners ranked them favorably is entirely consistent with reciprocal preferences.

Result 3. *Preference adjustments are largely reflecting reciprocal preferences in Info, as 73.68 percent of the adjustments are consistent with reciprocal preferences (while this fraction is only 20.69 percent in No-Info).*

Beyond establishing that information about others' preferences leads to higher instability, and that the preference changes are consistent with reciprocal preferences, our design allows us to pin down the underlying channels for these preference changes.

2.4 Mechanisms Underlying Reciprocal Preferences

In this section, we analyze the channels underlying reciprocal preferences using a theoretical model, which we then test empirically. In Section 2.4.1, we derive the optimal strategy of a proposer in a stylized version of the experimental *Info* condition and differentiate between belief-based and preference-based mechanisms. In Sections 2.4.2-2.4.4, we put the model's assumptions and implications to the empirical test.

2.4.1 Theoretical Framework

Two proposers (*he*) $p \in \{P, Q\}$ and two receivers (*she*) $r \in \{R, S\}$ participate in a simplified version of the matching market. The DA mechanism forms two teams, each with one proposer and one receiver, to play a PGG.[17] In this model, we allow proposers to be altruistic. Each proposer cares about their own direct (monetary) utility $u_p(\pi(c_p, c_r))$ which depends on the monetary payoff $\pi(c_p, c_r)$. The monetary payoff $\pi(c_p, c_r)$ is determined by both partners' contributions $c_{p,r} \in [0, 10]$. Selfish proposers ($a_p = 0$) follow a profit-maximizing strategy and free-ride ($c_p = 0$). Altruistic proposers ($a_p \geq 0$) care not only about their own direct (monetary) utility, but also about their matched partner's direct utility (u_r). The level of altruism $a_p \in [0, 1)$ towards the receiver depends on how likable the proposer perceives the receiver to be.

The core of our experimental treatment *Info* is that applicants learn how receivers rank them. We make two main assumptions about why this matters. Fist, we assume that the level of altruism is determined by the proposer's initial assessment of the receiver (l_p), and on how likable the receiver perceives him (l_r) to be. The level of altruism increases in l_p and l_r.[18] In other words, we assume that agents are more altruistic towards partners they like (Leider et al., 2009) and that "receiving information that another is attracted to you is a powerful determinant of liking" (R. M. Montoya & Horton, 2012). In our context, we assume that the receiver's rank is informative about l_r.[19]

Assumption 1. *Preference-based mechanism: The level of altruism (a_p) increases in l_r.*

Second, we assume that receivers contribute more to the PGG when being matched to a proposer they rank favorably. Proposers perfectly know the relation between the receivers' ranking (l_r) and their contributions' (c_r).

Assumption 2. *Belief-based mechanism: Receivers' contributions (c_r) increase in l_r.*

The direct (monetary) utility function $u_{p,r}$ is positive, monotonically increasing, continuous, and concave in the monetary payoff $\pi_{p,r}$ and has the same functional form for all agents. The adjusted utility of a proposer is given by:[20]

$$v_p = u_p(\pi_p(c_p, c_r)) + a_p(l_p, l_r) \cdot u_r(\pi_r(c_p, c_r))$$

These utility functions predict 1) how a proposer optimally selects his partner and 2)

[17] Section 2.2 offers a detailed description of the PGG and the DA mechanism. While participants in the experiment make discrete contribution choices, in Section 2.4.1 we assume that these are continuous.

[18] We assume that $l_{r,p}$ is a natural number.

[19] This is related to the idea of R. Dur (2009) that agent i's altruism towards another agent j depends the altruism of agent j towards agent i (which agent i infers from some action of agent j).

[20] The idea of direct (monetary) utility and adjusted utility is first described by Levine (1998).

how he decides about his contributions to the PGG. The timing of the model mirrors our experimental design in *Info*. First, proposers and receivers submit their preferences to the DA mechanism. At this point, the proposer has no information about l_r, his belief is the same for both receivers ($\hat{l_R} = \hat{l_S}$). This implies that proposers base their decision solely on l_p. Then, proposers learn the true preferences of both receivers (l_R, l_S). As proposers have (a priori) no information about l_r, being ranked first provides a weakly positive update about l_r while being ranked second presents a weakly negative update about l_r. Afterwards, proposers can adjust their ranking and play the PGG with their matched receiver. We solve the model by backward induction, first describing the contribution decisions before examining the implications for preference changes.

When matched with a receiver, a proposer optimizes by choosing his contribution to the PGG. Increasing the contribution level lowers his monetary outcome while raising the matched receiver's payout. The proposer's adjusted utility is maximized if the decrease in his marginal direct utility equals the increase in the matched receiver's marginal utility times the altruism factor towards her.

$$\max_{c_p} v_p : u_p(\pi_p(c_p, c_r)) + a_p(l_p, l_r) \cdot u_r(\pi_r(c_p, c_r)) \tag{2.1}$$

$$\frac{\partial v_p}{\partial c_p} = \underbrace{\frac{\partial u_p}{\partial c_p}}_{<0} + \underbrace{a_p(l_p, l_r) \cdot \frac{\partial u_r}{\partial c_p}}_{>0} = 0 \tag{2.2}$$

Following the optimization problem of the proposer, we give a short overview of the model's main predictions.[21] These proofs can be found in Appendix B.2.

Proposition 1. *An increase in l_r has a non-negative effect on the contribution of a proposer c_p.*

We assume that the level of altruism a_p (Assumption 1) and receiver's contribution $\hat{c_r}$ (Assumption 2) increase in l_r. If l_r increases, both channels then increase the proposer's contribution c_p in the case of an interior solution. First, as the level of altruism a_p increases, a proposer benefits more from the receiver's monetary payoff. Hence, the proposer's contribution c_p increases. Second, the higher contribution of the receiver decrease the proposer's marginal direct (monetary) utility and increase the receiver's marginal direct monetary utility. To equalize these marginal benefits (weighted by the

[21]The second order condition holds ($\frac{\partial^2 v_p}{\partial c_p^2} = \frac{\partial^2 u_p}{\partial {c_p}^2} + a_p \frac{\partial^2 u_r}{\partial {c_p}^2} < 0$).

altruism factor), the proposer increases his contribution.

Proposition 2. *A change in preferences for proposer P can only occur if a receiver R, whom proposer P initially ranked worse than receiver S, ranks him better than receiver S.*

If the proposer observes that he is ranked first by a receiver, he positively updates l_r. This change increases the proposer's adjusted utility of being matched with the receiver. Through a higher l_r (and hence a higher a_p and $\hat{c_r}$), the proposer both expects a higher monetary outcome for himself and cares more about the receiver. Both effects result in a higher contribution and lead to a higher utility for the proposer.

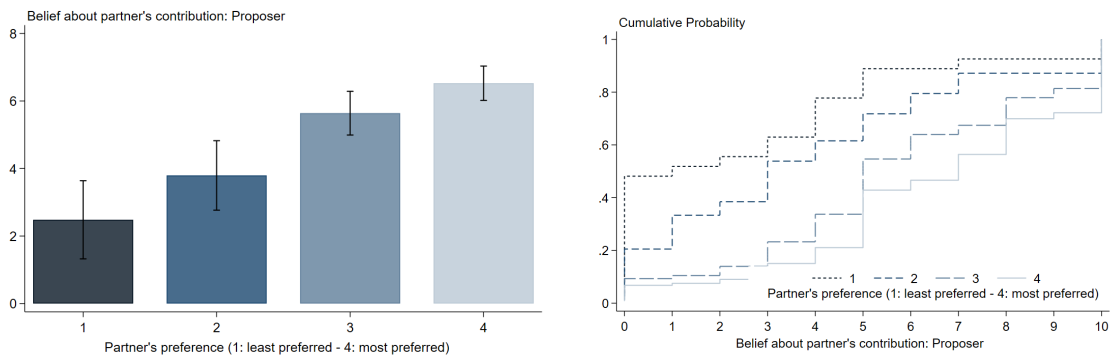
We can now derive the proposer's preferences over receivers and show why these preferences may change. A proposer ranks receivers based on his expected adjusted utility v_p of being matched with them if a strategy-proof mechanism is applied. His preference order can change upon learning how the receivers rank him. A positive update about l_r (weakly) increases the adjusted utility of being matched with a receiver. The reverse is true for a negative update. Therefore, a change of preferences can, for example, happen if the proposer initially ranked receiver R over receiver S, but then learns that he was ranked first by receiver S, and second by receiver R. This can, but need not, change the proposer's preference order. For an altruistic proposer ($a > 0$), these changes can be driven by preference-based and belief-based motives. For selfish proposers ($a = 0$), changes are entirely driven by beliefs about others' contributions. Selfish proposers will never contribute, but want to be matched to the highest contributing receiver.

Our model predicts preference changes consistent with reciprocal preferences (see Results 1 and 3). It highlights two channels for this behavior. First, participants change their preferences because they expect partners who like them to contribute more to the PGG. Preferences are interpreted as a signal about the match-specific value, and proposers change their preferences accordingly (belief-based). Second, a proposer may prefer to be matched with a receiver who liked them because he is more altruistic towards such a receiver (preference-based). Our results on the PGG behavior allow us to test whether preference-based or belief-based reasons explain the adjustments in *Info*, and how these adjustments translate into cooperative behavior.

2.4.2 Evidence for a Belief-Based Mechanism

We test the belief-based channel by analyzing (incentivized) beliefs of proposers about their matched receivers' contributions depending on how their partner ranked them. This means that we directly test our model's key Assumption 2 – that the receivers' preferences (l_r) are perceived as a signal about their contributions ($\hat{c}_r$). We first show that the receiver's preferences are indeed perceived as a signal about their contribution. We then demonstrate that these beliefs are accurate.

We find that proposers expect receivers who rank them better to contribute more to the PGG. Figure 2.3 shows this plotting beliefs over *Partner's preferences (1-4)*. This variable takes the value of four if the proposer was the matched receiver's most preferred choice, three if the participant was the second most preferred choice, and so on. Panel A shows that mean beliefs about the matched receiver's contribution increase with the receiver's preferences. Panel B illustrates this trend by presenting cumulative distribution functions. It shows, for example, that only 6.77 percent of proposers believe that their partner will contribute nothing when they were their partner's first choice. By comparison, 48.15 percent believed their partner will not contribute anything to the public good when they were their partner's least preferred choice.



A) Averages by *Partner's Preference* B) Distributions by *Partner's Preference*

Notes. This figure displays the beliefs of proposers in *Info* about the unconditional PGG contributions of their matched receiver by the preferences of the matched partner. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their matched partner, three if the participant was the second most preferred choice, and so on. Panel A shows averages, Panel B the cumulative distribution functions.

Figure 2.3: Beliefs about Receiver's PGG Contributions: Proposer in *Info*

Table 2.1 corroborates that proposers expect receivers who like to be matched with them to contribute more. The effects are sizable (Column 1), and remain so when controlling for the round and individual-level characteristics of the proposer (Column 2). Proposers expect matched receivers to contribute around 1.4 Taler (out of 10) more if they are ranked one place better on the receiver's preference list. This expectation is consistent with the notion that the expression of interest is "one cue to identify someone who is likely to act [...] cooperatively" (R. M. Montoya & Insko, 2008, p.478). Given such beliefs, a proposer may expect a change in their preference order to be payoff-maximizing if it results in being matched with a receiver who prefers them as a partner.

Table 2.1: PGG Behavior of Proposers in *Info*

	Belief Partner Contribution		Unconditional PGG Contribution		Avg. Conditional PGG Contribution	
	(1)	(2)	(3)	(4)	(5)	(6)
Partner's preference (1-4)	1.348***	1.382***	.771***	.794***	.415***	.416***
	[.977,1.720]	[.915,1.849]	[.350,1.193]	[.340,1.248]	[.128,.702]	[.147,.685]
Preference for partner (1-4)	-.073	-.059	.105	.146	.013	.026
	[-.437,.291]	[-.445,.328]	[-.219,.429]	[-.172,.463]	[-.182,.209]	[-.159,.212]
Round		.064		-.197***		-.119**
		[-.118,.245]		[-.338,-.055]		[-.214,-.023]
Loss Aversion		-.795***		-.710**		.040
		[-1.389,-.201]		[-1.408,-.012]		[-.429,.509]
Cognitive Ability (Raven's)		.500*		.337		-.159
		[-.037,1.037]		[-.347,1.021]		[-.557,.239]
Male		-.025		-.589		-.605
		[-1.434,1.383]		[-2.273,1.096]		[-1.581,.371]
Observations	285	285	285	285	285	285

Notes. OLS Regressions. *** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their matched partner, three if the participant was the second most preferred choice, and so on. *Preference for partner (1-4)* takes the value of four if the matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. *Round* is a count variable, indicating the number of the current round (Round 1-5). *Loss aversion* and *Cognitive ability* are calculated as detailed in Footnote 16, *Male* is an indicator taking the value of 1 if a participant indicated to identify as male.

Result 6. *Proposers expect a significantly higher unconditional contribution the more favorably the receiver ranks them ($p < 0.01$).*

The beliefs of proposers are in line with receivers' actual cooperation behavior. Figure B.1 displays that receivers contribute more to the PGG when matched with proposers they prefer. Table B.3 shows that each rank the matched proposer is up in the preference list leads to a 0.96 Taler higher contribution to the PGG in the pre-registered specification of Column 2, Table B.3.[22] Thus, proposers correctly expect receivers' preferences to influence their contribution decisions.

[22]Our design does not allow us to disentangle the underlying reasons for higher contributions by receivers. Still, our data is consistent with receivers (partly) contributing more when they like their partner, because they expect proposers they rank favorably list to contribute more to the PGG than proposers they rank less favorably (see Figure B.2 and Table B.7 in the Appendix). As receivers do not know that proposers learn their given rank, receivers believe that they can identify high-contributing proposers. In light of our findings that none of the personality questions predicts contributions in the PGG (see Table B.6), these beliefs turn out to be wrong.

In sum, we provide evidence for a belief-based mechanism underlying reciprocal preferences. We show that proposers rationally expect higher contributions from receivers who rank them favorably, which provides a rationale for the observed preferences adjustments.

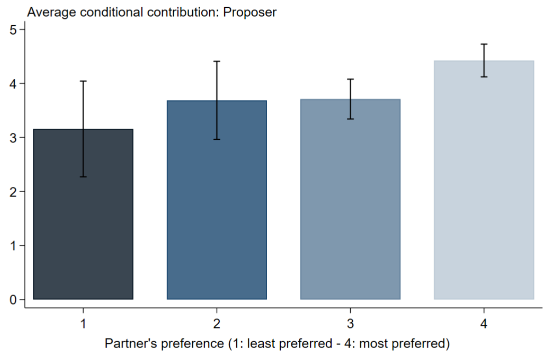
2.4.3 Evidence for a Preference-Based Mechanism

We test whether preference-based explanations play an additional role for reciprocal preferences by analyzing proposers' conditional contributions. Conditional contributions are independent from beliefs about the partners' contribution and, therefore, directly informative about the level of altruism (a_p). If proposers conditionally contribute more when interacting with a receiver who ranked them favorably, this implies higher altruism. Hence, we can directly test whether altruism is sensitive to the partner's preferences (l_r), which we presume by Assumption 1.

Proposers provide higher conditional contributions when matched to a receiver who ranks them favorably. Their average conditional contributions increase monotonically in the position on the receiver's preference list (see Figure 2.4, Panel A). Across the eleven conditional contribution decisions, they contribute around 0.4 Taler more for each spot they are ranked better (see Table 2.1). These averages mask an interesting heterogeneity, which we investigate in an exploratory analysis. Figure 2.4, Panel B shows that this difference in behavior is especially pronounced when facing higher contributions of the partner. The sub-figure plots the regression coefficient of the partner's preferences for each of the eleven contribution decisions, given the specification in Table 2.1, Column (6). For low contribution values of the receiver, the receiver's preferences do not strongly impact proposers' behavior. For example, proposers do not condition their contributions on whether a free-rider wants to be matched with them or not. However, receivers' preferences become an important determinant of proposers' conditional contributions when receivers make higher contributions. Proposers are then more altruistic towards receivers that indicate a preference to be matched with them.

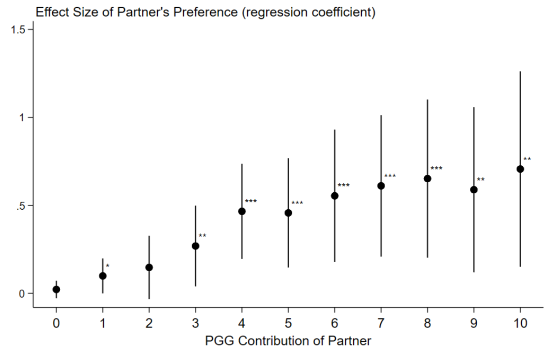
Result 7. *The conditional contributions of proposers are significantly higher when they interact with a receiver who ranks them favorably, especially for high levels of contributions by the receiver ($p < 0.01$).*

This provides evidence that preference-based explanations are important for the observed behavior. We document higher social preferences towards receivers who rank



Notes. This figure displays the average conditional contributions of proposers in *Info* by the preferences of matched receiver.

A) Averages by *Preference for Partner*



Notes. The figure plots the regression coefficients β_1 of the regressions $y_i = \beta_1 * Partner's\ preference + \beta_2 * Preference\ for\ partner + \beta_3 * t + \beta_4 * X_p$, corresponding to Table 2.1 with t indicating the round, and X_p as a vector consisting of gender, cognitive ability, and loss aversion. The outcome variables y_i is the conditional contributions of a proposer for any (unconditional) contribution $i \in 0, 10$ of the matched receiver. The vertical lines indicate the 95% confidence interval. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B) Coefficient Plot

Figure 2.4: Average Conditional PGG Contributions: Proposer in *Info*

the proposer favorably. Reciprocal preferences are therefore likely to stem both from preference-based and belief-based factors.[23]

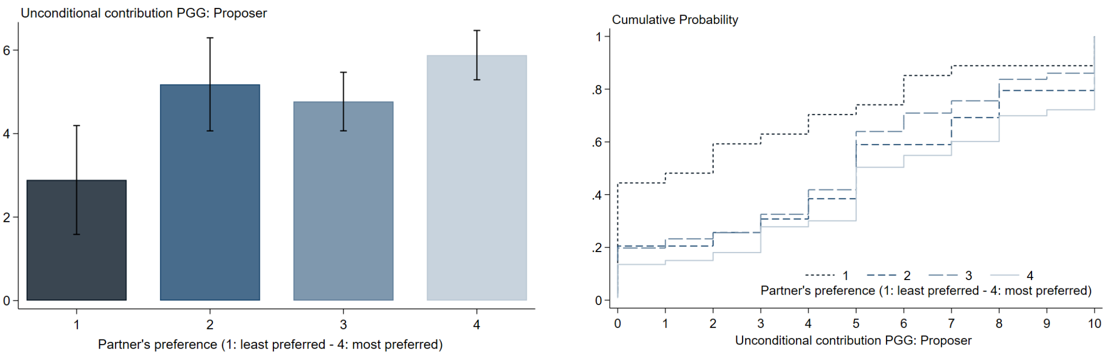
2.4.4 Unconditional Cooperation

Unconditional contribution decisions inform about the overall effect of reciprocal preferences in one-time simultaneous cooperation. Higher altruism (preference-based) leads to higher unconditional contributions. Higher beliefs about the contributions of the partner (belief-based) result in higher unconditional contributions by those willing to contribute more the more the other contributes (*i.e., conditional cooperation* as in Fischbacher, Gächter, & Fehr, 2001). The analysis of unconditional contributions directly tests Proposition 1.

On average, proposers contribute more to the PGG when interacting with a receiver who ranks them favorably. Table 2.1, Column 4 documents that proposers contribute around 0.8 Taler more when they are ranked one spot more favorably by their matched receiver. The partner's preferences (see l_r in the model) are more predictive of the actual contribution behavior of proposers than their own (initial) preference for the

[23]Table B.4 in the Appendix corroborates the robustness of these results in pre-registered analyses, showing that both mechanisms are specific to the information environment in *Info*.

partner (l_{p_r}). Figure 2.5 shows that unconditional contributions are especially low when interacting with a receiver who ranked them on the worst spot of their preference list.[24]



A) Averages by *Partner's preference*

B) Distributions by *Partner's preference*

Notes. This figure displays the unconditional contributions of proposers in *Info* by the preferences of the matched receiver. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their matched partner, three if the participant was the second most preferred choice, and so on. Panel A shows averages, Panel B the cumulative distribution functions.

Figure 2.5: Unconditional PGG Contributions: Proposer in *Info*

Result 8. *The unconditional contribution of proposers is significantly higher when they interact with a receiver who ranks them favorably ($p < 0.01$).*

Comparing social efficiency between treatments, we find that average (unconditional) cooperation and payoffs are higher in *Info* than in *No-Info*. While proposers in *No-Info* contribute on average 4.12 (out of 10) Taler to the PGG, contributions are around 25.7 percent higher in *Info* (5.18). On average, participants in *Info* contribute 0.96 Taler more to the PGG ($p = 0.039$; MWU), which translates into 0.48 Taler higher payoffs in *Info*. Accordingly, information about others' preferences increases average cooperation and payoffs.[25]

2.4.5 Gender Heterogeneity

Overall, male participants drive the differences in proposer's behavior depending on their partner's preference. In an exploratory regression analysis (Table 2.2), we show

[24] This is consistent with evidence from other domains that highlights the aversion to being ranked last, such as Kuziemko et al. (2014).

[25] Figure B.3 shows that the higher average payoff does not mask a substantial mean-variance trade-off. The treatment *Info* increases payoffs across the distribution.

that men's contribution decisions are significantly more influenced by their partner's preferences. This is true for both their unconditional (Column 2), and their conditional contributions (Column 3). In addition, men's beliefs (Column 1) about others' contributions are more responsive to their position on their partner's preference list.[26]

This gender heterogeneity raises interesting questions regarding how men and women react to rankings and evaluations. Previous research has found that women update more pessimistically than men when receiving negative feedback (Berlin & Dargnies, 2016). In addition, women attribute negative feedback to skill rather than to luck more often than men (Shastry, Shurchkov, & Xia, 2020), and react more strongly to likeability ratings based on their appearance (Gerhards & Kosfeld, 2020). In contrast, our results tend to suggest that men take the ranking more "personally" and react more strongly to it. This is consistent with previous findings recognizing women as being more ego-defensive (Möbius et al., 2022), and as having stronger internalized norms about giving, which leads to a lower elasticity of their altruism (Andreoni & Vesterlund, 2001). It is also in line with the finding of Barankay (2012) that feedback about performance rankings changes the behavior of men, but not of women.

Table 2.2: Gender Heterogeneity of Proposers in *Info*

	Belief Partner Contribution	Unconditional Contribution	Avg. Conditional Contribution
	(1)	(2)	(3)
Preference for partner (1-4)	-.086	.107	.006
	[-.462,.289]	[-.225,.438]	[-.174,.186]
Partner's preference (1-4)	1.004***	.253	.134
	[.457,1.551]	[-.225,.731]	[-.106,.373]
Partner's preference X Male	.798*	1.142***	.595**
	[-.037,1.634]	[.360,1.925]	[.123,1.068]
Male	-2.568*	-4.225***	-2.500**
	[-5.553,.418]	[-7.246,-1.205]	[-4.478,-.522]
Controls [Round + Individual]	Yes	Yes	Yes
Observations	285	285	285

Notes. OLS Regressions. *** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. *Preference for partner (1-4)* takes the value of four if the matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their matched partner, three if the participant was the second most preferred choice, and so on. *Male* is an indicator taking the value of 1 if a participant indicated to identify as male. *Partner's preferences x Male* takes the value of zero for observation with *Male=0*, and the value of *Partner's Preference X (1-4)* when *Male=1*. All columns control for *Round*, which is a count variable, indicating the number of the current round (Round 1-5), as well as *Loss aversion* and *Cognitive ability* that are calculated as detailed in Footnote 16.

[26]This leads to a lower average belief accuracy for men than for women (average deviation of 3.54 vs. 4.47, $p = 0.02$; MWU).

2.4.6 Similarity, Homophily, and Reciprocal Preferences

Perceived similarity influences behavior in various decisions (e.g., Eckel & Grossman, 2005; Y. Chen & Li, 2009; Hedegaard & Tyran, 2018), and has been shown to relate to interpersonal attraction (McWhirter & Jecker, 1967). Hence, the effect of partner's preferences' on their behavior may operate through a channel of homophily (McPherson, Smith-Lovin, & Cook, 2001; Currarini, Jackson, & Pin, 2009). If a proposer only has an imprecise signal about their similarity with the matched receiver based on five questions, the receiver's preferences (that are based on five different questions) may provide a signal about their similarity. Assuming common preferences to interact with a similar individual, the preference of the partner can be interpreted as information about their similarity.[27] So far, we have shown that information about others' preferences leads to more instability in matching markets through preferences adjustments (Section 2.3.1), that these adjustments are consistent with reciprocal preferences (Section 2.3.2), and that these adjustments likely stem from a combination of belief-based and preference-based factors (Sections 2.4.2 & 2.4.3). However, we have not yet established that the partner's preferences are not only similarity signals, but a fundamental determinant of behavior.

In Table 2.3, we provide evidence that partner preferences matter beyond being a signal for similarity. To do so, we add different measures of (objective) similarity to our main regression (Table 2.1). We see that our main effect persists when including these and conclude that similarity is unlikely to be the driver for our effects. We calculate similarity as the inverse of the average distance between the questionnaire responses of the matched partners (*Manhattan distance*). For example, the value is equal to 0 if one of the partners clearly affirmed each statement and the other clearly rejected all (i.e., the difference of their answers on the four-point Likert scale is maximal), and it is equal to 3 if they answered each question identically. First, the main coefficient of the partner's preferences remains constant when controlling for similarity based on all 15 questionnaire answers (Column 2). This implies that the partner's preferences do not fully operate by providing an accurate signal regarding similarity. Second, the main coefficient remains constant when including the similarity measure based on the five randomly selected questions for which the proposer has seen their partner's responses (Column 3). The positive and significant similarity coefficient implies that proposers

[27] Similar to Currarini and Mengel (2016), we find that similarity is an important predictor for partner choice in the PGG. The raw correlation between the rank given to a receiver and our basline measure for dissimilarity (Manhatten Distance) is 0.23, $p < 0.001$.

condition their contributions on whether their partner's responses match their own. At the same time, the similar main coefficients in Columns 2 and 3 imply that there is little additional signaling value in the preferences of the other agent. If there were, we would expect the main coefficient in Column 3 to be substantially higher than in Column 2. Third, the coefficient remains stable when we control for the similarity in answers across the five randomly selected questions to which the receiver has seen the proposer's answers. If preference were a signal about this similarity, this would again imply a lower main coefficient in Column 4 than in Column 1 or 2.

Table 2.3: Homophily and Unconditional Contributions of Proposers in *Info*

	Unconditional PGG Contribution (0-10)			
	(1)	(2)	(3)	(4)
Partner's preference (1-4)	.794***	.704***	.812***	.794***
	[.340,1.248]	[.252,1.155]	[.371,1.254]	[.344,1.245]
Preference for partner (1-4)	.146	.044	.123	.139
	[-.172,.463]	[-.313,.402]	[-.195,.441]	[-.181,.458]
Similarity Answers (0-3) [Manhatten]		1.357		
		[-.563,3.278]		
Similarity of Shown Answers (0-3) [Manhatten]			1.041**	
			[.115,1.968]	
Similarity of Receiver's Answers (0-3) [Manhatten]				.862
				[-.293,2.017]
Controls [Round + Individual]	Yes	Yes	Yes	Yes
Observations	285	285	285	285

Notes. OLS Regressions. *** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their matched partner, three if the participant was the second most preferred choice, and so on. *Preference for partner (1-4)* takes the value of four if the matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. similarity is the inverse of the average distance between the questionnaire responses of the matched partners (Manhattan distance). *Similarity Answers (0-3)* is calculated based on all 15 questionnaire items, *Similarity of Shown Answers (0-3)* is based on the five questions the proposer saw the partner's answers for, *Similarity of Receiver's Answers (0-3)* is based on the five questions the receiver saw the partner's answers for. All columns control for *Round*, which is a count variable, indicating the number of the current round (Round 1-5), the indicator *Male* taking the value of 1 if a participant indicated to identify as male, as well as *Loss aversion* and *Cognitive ability* that are calculated as detailed in Footnote 16.

2.5 Conclusion

This paper shows that reciprocal preferences represent a powerful source of instability in matching markets. First, we demonstrate that reciprocal preferences exist – that is, that participants *like to be liked*. When participants learn the preferences of their potential partners, they adjust their preferences and rank more favorably those who would like to be matched with them. Second, we show that these changes substantially increase the number of unstable matchings. Third, we investigate the underlying motives of reciprocal preferences and find evidence for both belief-based and preference-based mechanisms. On the one hand, proposers expect receivers who like them to contribute more to the PGG. This provides a profit-maximizing rationale for preference adjustment due to changes in beliefs. On the other hand, proposers are more altruistic towards receivers who like to be matched with them. This supports a preference-based rationale for reciprocal preferences.

The PGG reflects the cooperative nature of many matching markets. In matching markets, not only are relationships formed without the coordinating function of prices, but also within the relationships there are non-contractible elements. Insofar as these elements relate to effort provision and commitment, cooperation plays a crucial role in these relationships. Consider a university that wants to hire an enthusiastic job market candidate, and (in turn) a candidate who also wishes to receive support from the department. Both choosing a cooperative partner *and* being in a relationship where one wants to be cooperative oneself is key in such a setting, where decisive aspects cannot be contracted upon. The PGG allows us to investigate both of these channels.

Notwithstanding, our stylized experimental setting does not capture all aspects of the preference-based foundation of reciprocal preferences. Real interactions put more weight on psychological mechanisms, such as the non-pecuniary disutility of working with someone who does not like you. Therefore, investigating reciprocal preferences in inter-personal coordination tasks constitutes an avenue for future research. Compared to our experimental design, which likely provides a lower bound for the effect of reciprocal preferences, the effects could be even more pronounced when individuals expect a personal interaction.

Our results contribute to the understanding of matching markets, cooperative behavior, and effective team formation in organizations. First, our results can help to design matching markets more efficiently. It is necessary to understand why matching markets sometimes fail to reach their full potential. Opitz and Schwaiger (2023b) theoretically show that reciprocal preferences can be a source of instability. Evidence from real-world matching markets suggests that reciprocal preferences play an important role. Nevertheless, observational data does not allow for teasing apart reciprocal preferences, uncertainty, and other potential reasons for market failures. This paper establishes the empirical relevance of reciprocal preferences and thus highlights the importance of information design in matching markets. While learning about others' real-world preferences might sometimes be more subtle than in the experiment, already observing the final matching can lead to updates about other participants' preferences and result in instability. Our sizeable effects suggest that reciprocal preferences also play an essential role in slightly different information environments. Understanding the importance of reciprocal preferences helps to reconcile strategic modifications of the theoretically efficient mechanism by participants (e.g., by offering early admission (Avery & Levin, 2010), by making admission decisions contingent on others' preferences (U. Dur et al., 2022), or by introducing preference signaling devices (S. Lee & Niederle, 2015)). In addition, it helps to design mechanisms that accommodate agents' reciprocal preferences (Opitz & Schwaiger, 2023a). Indirectly, we also provide evidence that agents may misrepresent their preferences in strategy-proof environments because they believe higher rankings to be rewarded by agents on the other side of the market (see Hassidim, Romm, & Shorrer, 2021), as is the case in many economic environments – including our experimental one.

Second, we enhance understandings of social preferences and social proximity. Previous research shows that we treat those close to us more favorably, without being able to differentiate between our liking, being liked, and similarity (Leider et al., 2009). By isolating the role of being liked, we provide evidence that giving in a relationship depends on not only our own preferences, but also others' preferences. These findings are consistent with literature outside economics that emphasizes the wish to be liked as a universal desire (Baumeister & Leary, 1995) with neural underpinnings (Davey et al., 2010), and the susceptibility of our own interpersonal preferences to the preferences of others (R. M. Montoya & Horton, 2012, 2014). We demonstrate that this susceptibility implies that interpersonal preferences are another currency of reciprocity, expanding previous findings on which type of gifts can lead to productivity gains (e.g., Kube, Maréchal, & Puppe, 2012). Hence, we link interpersonal preferences to organizational implications for motivating workers.

Third, our findings on the relevance of reciprocal preferences have broader organizational implications for team formation and teamwork. Organizational processes and production steps require voluntary cooperation to achieve optimal results (Deversi, Kocher, & Schwieren, 2020). We show that being liked can be necessary for cooperation. Previous literature has established that self-selected teams display homophily in their traits and networks, leading to higher satisfaction and effort (R. Chen & Gong, 2018; Boss et al., 2023). We provide a foundation for these results by highlighting greater cooperation when collaborating with a partner who likes you. We show that even in a stylized setting without personal interactions, we observe homophily in sorting, and higher cooperation among those matched with partners who like them.

3

Interpersonal Preferences and Team Performance

The Role of Liking in Complex Problem Solving

3.1 Introduction

Teamwork is increasingly important in modern organizations. In addition, today's work tasks are more complex, more abstract, and more non-routine (Autor, Levy, & Murnane, 2003; Autor & Price, 2013). These tasks typically require workers to apply problem solving skills, rely on intuition, and be creative. Especially in complex tasks, teams have the potential to perform more efficiently than individuals due to skill complementarities (e.g., Ichniowski & Shaw, 2013). However, for these to materialize, teams must assign responsibilities, collaborate, communicate effectively, share information, and arrive at good solutions together (Deming, 2017). Thus, even "a team of experts does not necessarily make an expert team" (Salas, Reyes, & McDaniel, 2018).

Therefore, it is critical to understand how to make teams most effective when faced with complex problem solving. This requires an analysis of the environment in which teams operate, the skill composition of team members, as well as the collaboration within a team. Previous literature documents positive effects of incentive structures such as bonuses and tournaments on team performance (Englmaier et al., 2023a, 2023b), demonstrates the value of leadership (Englmaier et al., 2021), and studies optimal team composition based on individual characteristics in non-routine tasks (Hoogendoorn, Oosterbeek, & Van Praag, 2013; Hoogendoorn, Parker, & Van Praag, 2017; Hardt, Mayer, & Rincke, 2023). However, it remains challenging to predict how well a given team collaborates. Realizing the full potential of teams in non-routine tasks requires a deeper understanding of the determinants of successful team collaboration.

Interpersonal preferences of team members can play a crucial role for collaboration and team performance. Knowing that one is in a team with others who like them potentially increases satisfaction, facilitates helping behavior, or provides psychological safety to voice concerns and express opinions openly. All of these may contribute to a better *team spirit*[1], which is considered to be essential for team success and can explain why we often prefer to interact with those who also want to interact with us (Opitz & Schwaiger, 2023a). In this paper, I examine whether teams perform better when team members like each other.

I conduct a laboratory experiment to test whether interpersonal preferences causally affect team performance in a non-routine task. The laboratory setting allows me to infer participants' interpersonal preferences and to manipulate information structures. My underlying hypothesis is that teams work better when the team members like each other. There can be two reasons for this. On the one hand, people who like each other may be more successful because of sorting. Given that perceived similarity has been shown to relate to interpersonal attraction (McWhirter & Jecker, 1967; R. M. Montoya, Horton, & Kirchner, 2008), those who like each other may be more similar and therefore exert higher effort (e.g., Y. Chen & Li, 2009) or communicate more effectively. On the other hand, the feeling of being liked may causally change behavior. For example, people may adjust their behavior when they learn that their team members like them. They may provide higher effort, feel safer on the team, and ultimately become more productive and satisfied. In the experiment, I not only test whether those who like each other work better together, but also disentangle the underlying mechanisms by

[1]*Team spirit* is defined as *The willingness [...] to work together and support[...] each other as part of a team* (New Oxford American Dictionary, 3rd ed.).

selectively revealing information about others' preferences.

The experiment consists of a team formation process and a non-routine team task that participants perform under one of two information structures. During the team formation process, participants indicate with whom they would like to interact in a payoff-irrelevant situation. They rank their potential partners on the basis of short personality profiles. Participants then perform a non-routine team task with a randomly assigned partner from those ranked before. This task requires participants to find a numerical solution to an abstract problem, often referred to as a *Guesstimation* task (Morgan, Neckermann, & Sisak, 2021).[2] Many companies use such tasks in job interviews because they demonstrate critical problem solving abilities and require coordination and communication –skills that have been identified to be key for the success in 21st century workplaces (Binkley et al., 2012; Aerlebaeck & Albarracin, 2019). In the baseline condition (*No-Info*), participants never learn how their partner ranked them. In the treatment condition (*Info*), participants receive information on how they are ranked by their matched partner before performing the *Guesstimation* task jointly. This allows me to compare whether teams where partners learn that they like each other perform better than teams who learn that the partners do not like each other. By contrasting behavior with the *No-Info* condition, I rule out the possibility that performance differences are due to people who like each other having *something* in common that might enhance performance. If this were the underlying mechanism, it would also operate when the partner's preferences are unknown. In this way, I establish that the disclosure of interpersonal preferences is the underlying mechanism.

I find that interpersonal preferences matter for performance in complex problem solving. While teams in which partners like each other perform similarly to teams in which partners dislike each other, teams in which one partner likes the other more than the other (*dissimilar liking*) perform best. Changes in collaborative behavior upon learning the preferences of the team partner are the causal mechanism underlying these performance differences. Participants do not anticipate this. Before the task, they expect to be more successful when being in a team where partners like each other due to higher effort provision. I present suggestive evidence that the performance differences are driven by different communication patterns. In retrospect, participants evaluate the collaboration in a similar way depending on whether they were in a team where

[2] A classic example of a *Guesstimation* task (also referred to as a Fermi problem) is the question of the number of piano tuners in Chicago. The challenging thing about such problems is that individuals neither have direct empirical values from a similar problem, nor do they have the necessary data with which they could directly make a calculation.

the partners liked each other, disliked each other, or had dissimilar preferences.

This paper contributes to three strands of literature. First, I add to an organizational literature on the self-selection of teams. Traditionally, management has been responsible for the composition of teams in firms. But there is an increasing development towards giving workers more flexibility in choosing their team partners –a situation that is the norm in other environments such as academic institutions or entrepreneurship.[3] Performance in self-selected teams can be higher because workers have better information on how to form teams effectively. On the other hand, workers may also form teams based on personal preferences that are not conducive to performance. Bandiera, Barankay, and Rasul (2013) illustrates this trade-off in a manual task by showing that team based incentives induce workers to team up with others of similar ability and forgoing the non-pecuniary benefit of working with their (lower-ability) friends. Through a structural model, Allocca (2023) shows higher performance of self-selected teams in a scientific organization. This is in line with findings of Boss et al. (2023), demonstrating the benefit of freely choosing team partners for a given pitch deck presentation task. Similarly, Fischer, Rilke, and Yurtoglu (2023) shows higher performance of self-selected teams in a task environment where the abilities of all team members play a significant role in the team production function. I contribute to this literature by showing that those who like each other do not necessarily perform better in complex problem solving, despite the expectation of higher team effort and better performance. This underscores the importance of understanding the dimensions on which individuals self-select into teams when assessing the effectiveness of giving them the freedom to do so.

Second, I relate to to the literature studying interpersonal preferences in teamwork. Gächter, Starmer, and Tufano (2022) demonstrates the importance of social relationships for team production. They find a strong positive association between group cohesion and performance in weak-link coordination games. This is consistent with evidence that social proximity can lead to higher levels of prosocial behavior, which is an important prerequisite for successful cooperation (Leider et al., 2009; Goeree et al., 2010). Opitz and Schwaiger (2023a) use an approach similar to this paper in which interpersonal preferences are based on questionnaire responses within the experiment instead of relying on real-world friendship networks. They find higher cooperation in a Public Goods Game when participants are liked by their partner, and point to in-

[3] Some companies give employers complete flexibility in choosing projects and teams (e.g., Zappos), others allow employees to reshuffle teams in regular intervals (e.g., Microsoft), and still others have a long history of encouraging side-project time in self-selected teams (e.g., Google).

creased altruism and higher beliefs about the other's contributions as the underlying pathways.

Third, the paper advances the literature on team performance in non-routine tasks. There is increasing interest in understanding the production function in complex work tasks through experimental approaches. Research questions range from the optimal team composition (Hoogendoorn, Oosterbeek, & Van Praag, 2013; Hoogendoorn, Parker, & Van Praag, 2017; Hardt, Mayer, & Rincke, 2023), to identifying team players (Weidmann & Deming, 2021), to effective incentive and governance structures (Englmaier et al., 2021; Morgan, Neckermann, & Sisak, 2021; Englmaier et al., 2023a, 2023b). The experimental approaches range from field experiments on entrepreneurship education programs and escape room games (e.g., Hoogendoorn, Oosterbeek, & Van Praag, 2013; Hoogendoorn, Parker, & Van Praag, 2017; Englmaier et al., 2023a) to laboratory experiments (Morgan, Neckermann, & Sisak, 2021) and online experiments (Hardt, Mayer, & Rincke, 2023). In this paper, I shed light on how interpersonal preferences affect collaboration of teams in a laboratory setting by employing a *Guesstimation* task, using chat communication.

The paper is structured as follows: Section 3.2 presents the experimental design, Section 3.3 outlines my hypotheses and results on interpersonal preferences in non-routine teamwork. Finally, I discuss and conclude in Section 3.4.

3.2 Experimental Design

Overview. The experiment consists of two parts. First, a team formation process. Second, a non-routine team task. During the team formation process, participants indicate with whom they want to interact in a payoff-irrelevant situation. This means that participants rank their four potential partners according to the desirability of interacting with them without this decision being tied to any potential payoff. This is how I operationalize *liking*. These interpersonal preferences are based on short profiles of the potential partners. The incentivized non-routine task is played with one randomly matched partner from those ranked. I analyze behavior under two information structures. In the baseline condition (*No-Info*), participants never know how much their partner likes them. In the treatment condition (*Info*), participants receive the information before the non-routine team task. I investigate whether performance in the non-routine task differs depending on how much team partners like each other, and I analyze different mechanisms of how this may affect team performance. The timeline of the experiment is visualized in Figure C.1.1.

Team Formation. I form teams for the *Guesstimation* task in order to analyze whether performance differs depending on whether team partners like each other or not. I face two challenges in doing so. The first is to operationalize *liking*. Ideally, one needs to elicit incentive-compatible preferences about the desirability of interacting with each other, without these preferences being linked to the *Guesstimation* task. The second challenge is to create teams with different preference constellations. I intend to compare teams where both partners like each other, to teams where only one partner likes the other, to teams where both partners dislike the other. This can be achieved through randomization. By creating a payoff-irrelevant interaction, I solve both challenges.

Participants express their preferences for potential interaction partners based on seeing questionnaire responses of them. At the beginning of the experiment, participants complete a personality questionnaire. This contains 15 statements on personality traits, preferred leisure activities, and societal opinions that participants answer on a 4-point Likert scale (see Appendix C.1.2). Intuitively, participants get an impression about each other based on seeing the responses to the questionnaire. They are then asked to rank their potential partners according to the desirability to interact with them, which I use as a proxy for *liking*. Participants have no further information about the payoff-irrelevant interaction the ranking refers to, so they do not know what the *interaction* will be about.[4] Their preferences therefore directly reflect which of their potential partners they enjoy interacting with the most and least.

Practically, participants are randomly assigned to groups of 8 players, whereby each group is evenly split into two *market sides*. Then the questionnaire responses to five randomly selected questions are shared between the participants across market sides. That is, each market side receives the answers to these questions from the four players on the other market side. Participants then rank the players based on their perception of the desirability of interacting with them.[5] The preference submission takes place via a strategy-proof mechanism (Abdulkadiroğlu & Sönmez, 1998). It is in the best

[4]Ultimately, participants are asked to create three statements that they believe half of the student population would agree with and half would disagree with during this interaction.

[5]Participants have no additional information about each other such as gender, age, or sociodemographic characteristics. They only see the answers to the five questions.

interest of the participants to submit their true preferences.[6] This mechanism then determines the teams for the interaction.

Non-routine Team Task. The non-routine team task is a *Guesstimation* problem.[7] The goal is to estimate an unknown and highly unusual quantity by connecting a series of known or easily estimated quantities in a logical but non-routine way. Participants are paid for the accuracy of their response (Morgan, Neckermann, & Sisak, 2021). They are allowed to use a calculator, but are not given the possibility to search for additional information on the internet. To familiarize participants with the task and to get a measure of individual ability, they first perform one *Guesstimation* task on their own.[8]

Participants work on the team task with one randomly selected team partner. The allocation mechanism randomly pairs individuals from each market side within the group of 8 players that ranked each other during the team formation stage. Thus, I create random teams out of players who indicated how much they *like* each other. Teams have 8 minutes to solve the joint task, and communicate via chat messages. For a team's answer to count, both team partners must enter the same solution. This requires close collaboration between the two partners.

Treatments. I assess whether interpersonal preferences affect performance in the non-routine team task. Participants perform the task under one of two information structures. They either learn how much their team partner likes them or they do not. That is, in one condition, participants learn the rank their partner placed them on for the interaction, in the other condition, they do not.

The treatment variation takes place precisely at the time of learning who the team partner is. In the experimental condition *No-Info*, participants learn who their team partner is. This includes a reminder of how they ranked their team partner, and how this partner answered the randomly chosen questions. In the experimental condition *Info*, participants additionally learn how their assigned partner ranked them. In this way, I reflect situations in which team partners like or dislike each other for reasons orthogonal to the specific task they perform (here the *Guesstimation* task). Treatment

[6] Participants are also informed that the indication of one's true preferences increases the likelihood of performing the interaction with this person. I use a random serial dictatorship mechanism. Thereby, participants are sorted and randomly listed in order. The first in line is then assigned their favorite partner, then it is the second's turn, and so on until the four two-player teams are formed in each group of 8 participants.

[7] See Appendix C.1.3 for the set of problems.

[8] For the individual task, participants are given 5 minutes and can earn up to 5 Euro.

is assigned at the group level. This means that either both team partners know about the other's preference, or neither partner knows.

Underlying Channels. I investigate the channels through which *liking* may affect task performance by capturing participants' ex-ante beliefs about their collaboration as well as their ex-post evaluation of it. Before the task, I assess whether participants expect to be more successful and enjoy the task more when they are in a team with partners who like each other. I distinguish between two motives for why participants may show a preference for working with someone they like and/or who likes them (Opitz & Schwaiger, 2023a). On the one hand, a belief-based motive related to higher expected team performance. On the other hand, a preference-based motive about higher expected utility from doing the task itself. After the task, I ask participants about their collaboration experience and their willingness to continue working with their current partner, capturing both their success beliefs and their enjoyment of the task.

I elicit four ex-ante beliefs about the upcoming *Guesstimation* task once participants know who their partner is. First, they are asked about their motivation to provide high effort in the upcoming task on a scale from 0 to 10. Second, participants indicate their belief about their partner's effort. This belief is incentivized through the partner's self-reported effort on the previous question. If the respondent's assessment matches their partner's self-reported effort, they receive 2 Euro. If the assessment is not equal to the partner's report, they receive no payment. Third, participants provide their team performance beliefs. For this, I endow participants with 1 Euro. They can bet any amount (in increments of 10 Cents) on their own team's performance and keep the rest. If their team happens to be in the 25% best performing teams, the amount invested is quadrupled and paid out. If not, the investment is lost. In this way, they indirectly indicate their expected probability that their team will be among the 25% best performing teams.[9] Fourth, participants indicate their willingness to pay to perform an extra round of the task with fixed pay alone compared to with their partner.[10] This allows me to measure their expected task utility, the utility of doing the task with their partner regardless of any payoffs.

After the completion of the *Guesstimation* task and before any performance feedback,

[9] I either pay out the beliefs about partner's effort or the beliefs about team performance with equal probability. I inform participants before the *Guesstimation* task which belief is drawn to be payoff-relevant.

[10] The multiple price list procedure to elicit the willingness to pay is displayed on Screen 10 in Appendix C.3.

participants assess their collaboration experience on two different dimensions. First, they rate different aspects of their collaboration experience (see Appendix C.1.2 for the questionnaire). Second, they decide whether they would like to perform an extra round of the task with performance-contingent payment with their team partner or alone. This binary decision captures both their (relative) performance beliefs as well as their task utility from doing the task with their team partner.

Of the two choices of whether to work alone or with your partner on an extra task, at most one will be implemented. There are three possible implementation procedures. First, participants do not do another round of the task (implemented with 90 percent probability). Second, participants do another round of the *Guesstimation* task with a fixed payoff. This is the scenario to which the ex-ante measure for expected "task utility" refers (5 percent probability). Third, participants do another round of the *Guesstimation* task with performance-contingent payoff. This is the scenario the ex-post measure for "task continuation" refers to (5 percent probability). In this way, I elicit incentive-compatible preferences over teamwork, but make sure that most participants do not have to do an extra round of the task. Given that there is an extra task, I follow the implementation procedure of Hardt, Mayer, and Rincke (2023): There is a 50 percent chance that participants get to work alone irrespective or their indicated preferences. Likewise, there is a 50 percent chance that their choice will be implemented. In particular, participants get to work together if both partners stated a preference for doing so.[11]

Experimental Procedures. The experiment was conducted at the Munich Experimental Laboratory for Economic and Social Sciences (MELESSA) in July 2023, with physical participation in the laboratory. In total, 240 student participants participated in the experiment across 13 sessions. The participants were recruited using the online system ORSEE (Greiner, 2015). The experiment was programmed with the software oTree (D. Chen, Schonger, & Wickens, 2016). Treatment was randomized within-session at the group level. On average, participants earned 19 Euro (including a show-up fee of 10 Euro). The experiment lasted about 55 minutes.

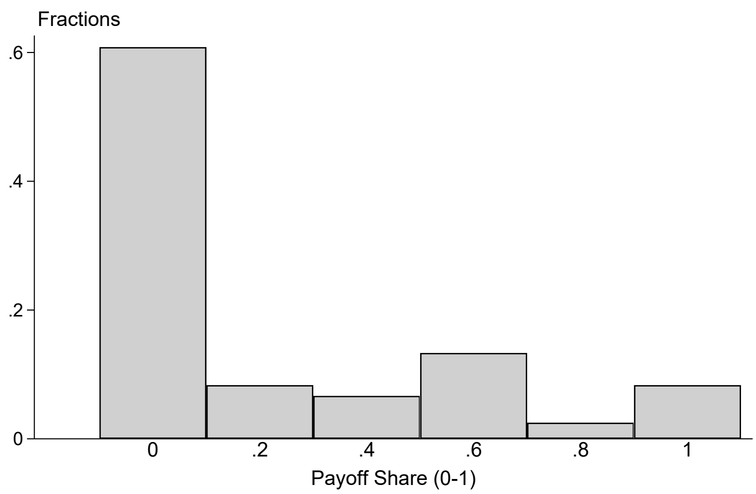
[11]This requires first randomly selecting one of the rows in the multiple price list as the relevant one for the (ex-ante) assessment of expected task utility.

3.3 Main Results

3.3.1 Incentives and Task Performance

Participants could earn up to 10 Euro in the (collaborative) *Guesstimation* task. Following the incentive scheme of Morgan, Neckermann, and Sisak (2021), participants received the full payment if their answer deviated less than 10% from the true value. For deviations of up to 20% (40%/ 60%/ 80%) they received 8 Euro (6 Euro/ 4 Euro/ 2 Euro). For larger deviations or no response, they did not receive any payment.

Out of the 120 teams in the experiment, 72 submitted a valid answer within the given time. Those who did not submit a valid answer either failed to submit an answer within the time limit (33 teams), or gave different answers from each other (15 teams). From those who submitted a valid answer, around 65 percent achieved a positive payoff, while the rest of the teams submitted a solution that deviated by more than 80% from the true value. On average, participants earned 2.23 Euro from the collaborative *Guesstimation* task, Figure 3.1 shows the payoff distribution. The variable *Payoff Share* refers to the incentive scheme participants faced. For example, a value of 0.4 means that the participant received 4 out of the possible 10 Euro and implies that the answer deviated between 40% and 60% from the true value. Figure C.3 in Appendix C.2.8 shows that performance across the five *Guesstimation* tasks was similar, ranging from averages of 0.19 to 0.24 in the payoff share.



Notes. This figure displays the distributions of the payoff share in the collaborative *Guesstimation* task based on all 120 teams allocated to either the *Info* or *No-Info* treatment.

Figure 3.1: Payoff across Treatments

3.3.2 Estimation Strategy

I perform two different types of analyses to answer whether people who like each other perform better in complex problem solving and why. The first type of analysis examines behavior within the *Info* treatment. Conceptually, I compare teams where the partners like each other, to teams where the partners dislike each other, to teams where one partner likes the other more. This allows me to make general statements about which type of team performs best. However, this analysis does not disentangle whether the underlying mechanism is sorting or a causal effect of feeling liked by the team partner.

I analyze underlying mechanisms by comparing behavior across the treatments *Info* and *No-Info* in the second type of analysis. That is, I compare behavior of those who like each other (or not) across *Info* and *No-Info*. Thereby, I hold the underlying preferences constant, and isolate the effect of "knowing these preferences". For example, I can compare a team where both partners ranked each other first and knew it (in *Info*) with a team where both partners ranked each other first but did not know it (in *No-Info)*. If people who like each other are, for example, more similar and this per se facilitates communication, these teams perform better even when participants do not know their partner's preferences (in *No-Info*). In contrast, finding that interpersonal preferences affect performance in *Info* but not in *No-Info* provides strong evidence for a causal effect of feeling liked and related changes in collaborative behavior.

For both types of analyses, I run two sets of regressions. In one set, I estimate the effect of the sum of both partners' preferences on the outcome. Preferences are a reverse-coded measure of the ranks they assign to each other. Setting a partner on the first and most favorable rank is converted to the highest possible preference, setting a partner on the last rank is converted to the lowest possible preference. In the other set, I add a measure of how dissimilar partners' preferences are. This allows me to analyze, for example, whether it makes a difference if both partners rank each other second, or if one partner ranks the other first and one partner ranks the other third in their preference list.

3.3.3 Liking and Task Performance

Liking within Treatment *Info*

Teams where partners like each other do not perform better than teams where partners dislike each other. Column (1) of Table 3.1 shows a negative but statistically insignificant impact of *joint liking* on performance. *Joint Liking* is defined as the sum of both partners' preference rankings. In Column (2) I add controls for team demographics and individual performance. It confirms a statistically insignificant but economically significant negative impact of about 18.5 percent on payoffs for each rank that one of the two partners ranked the other more favorably.

Teams with *dissimilar liking* perform better. I define *dissimilar liking* as the absolute difference between partners' preference rankings. The pre-registered analyses in Columns (3) & (4) indicate that teams where one partner likes the other more than vice versa are more successful. This also holds in Column (5) when adding fixed effects for the different *Guesstimation* problems. The coefficient for *joint liking* remains in the same magnitude when including *dissimilar liking* and only reaches marginal significance in Column (4). I summarize the main findings in Results 1 and 2.

Table 3.1: Task Performance: Guesstimation [*Info*]

	Payoff Share [0-1]				
	(1)	(2)	(3)	(4)	(5)
Joint Liking (2-8)	-.032	-.037	-.035	-.043*	-.039
	[-.078,.013]	[-.085,.011]	[-.080,.010]	[-.089,.004]	[-.088,.010]
Dissimilarity Liking (0-3)			.063	.094**	.121**
			[-.018,.143]	[.008,.180]	[.024,.219]
Mean Dep. Var.	.20	.20	.20	.20	.20
Team Controls	No	Yes	No	Yes	Yes
Performance Controls	No	Yes	No	Yes	Yes
Problem FE	No	No	No	No	Yes
N	60	60	60	60	60

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is the fraction of the maximum possible payoff that teams achieved. The values in square brackets represent the 95% confidence intervals. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants face.

Result 1. *Teams where partners like each other do not perform better than teams where partners dislike each other in Info.*

Result 2. *Teams where partners have dissimilar preferences perform significantly better in Info.*

In the next subsection, I analyze whether this pattern is driven by sorting. I show that partners who like (or dislike) one another do not perform differently per se. This is despite the fact that I observe homophily in the sense that those who are similar also like each other more (McPherson, Smith-Lovin, & Cook, 2001). The more similar a potential partner's answers are to one's own answers, the more favorable this partner is ranked ($r = 0.27$. $p < 0.001$; Pearson's correlation coefficient). This implies that teams where both partners like each other are, on average, most similar in terms of their questionnaire responses. Theoretically, this can either improve performance by making it easier to communicate and share information, or it can be detrimental if similar partners have less complementary skills and information. Empirically, I show that sorting does not affect performance at large. Instead, the patterns in Results 1 & 2 stem from knowing how much their partner likes them.

Disentangling Selection and the Feeling of Being Liked

I show that the knowledge about the partner's preferences changes performance through comparing behavior across the treatments *Info* and *No-Info*. Table 3.2 corroborates the findings from analyzing the treatment *Info* in isolation. First, it shows that the preferences of team partners only matter in the *Info* treatment. If participants who like each other naturally worked better (or worse) together because they have similar communication patterns, thinking styles, or skills, this would also show up in *No-Info*.[12] Second, it confirms that participants who like each other do not perform better when knowing this information. The effect is consistently negative and statistically significant across Columns (1)-(3), showing that learning about each others' favorable preferences is detrimental to performance. Third, also this analysis supports that teams with dissimilar liking perform better in *Info*. Coefficients in Columns (2) & (3) are economically sizeable and statistically significant at the 10% level.

Result 3. *Performance differences between teams with different interpersonal preferences are due to changes in behavior upon learning how partners like each other.*

[12] In Table C.1, I also analyze the treatment *No-Info* in isolation and consistently shows that this is not the case. Additionally, I show in Table C.14 that questionnaire responses and performance are largely uncorrelated. This means that participants cannot (objectively) identify who is good at the task individually (Column 1), whether someone is a good team partner in the task (Column 2), or whether someone is a good team partner given the available information about each others preferences (Column 3).

Table 3.2: Task Performance: Guesstimation [across Treatments]

	Payoff Share (0-1)		
	(1)	(2)	(3)
Joint Liking (2-8)	.034	.035	.036
	[-.012,.081]	[-.011,.081]	[-.011,.083]
Joint Liking X Info	-.068*	-.073**	-.072**
	[-.137,.001]	[-.142,-.005]	[-.143,-.001]
Info	.303*	.197	.170
	[-.050,.656]	[-.170,.564]	[-.218,.558]
Dissimilarity Liking (0-3)		-.040	-.038
		[-.145,.064]	[-.146,.070]
Dissimilarity Liking X Info		.124*	.134*
		[-.015,.263]	[-.011,.279]
Mean Dep. Var.	.23	.23	.23
Team Controls	Yes	Yes	Yes
Performance Controls	Yes	Yes	Yes
Problem FE	No	No	Yes
N	120	120	120

Notes. OLS Regressions.*** $p < 0.01$, **$p < 0.05$, * $p < 0.1$. The outcome variable is the fraction of the maximum possible payoff that teams achieved. The values in square brackets represent the 95% confidence intervals. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Joint Liking X Info* interacts this variable with a treatment indicator for *Info*. *Dissimilarity Liking* is the absolute difference between both partners' preferences, *Joint Liking X Info* captures the interaction with the treatment *Info*. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants face.

I confirm these results in a robustness check where I correct for idiosyncratic mistakes when entering the final response. There are 12 teams that arguably coordinated on a common response, but one of the team members made a mistake when entering the solution. These mistakes are either typos or formatting errors. Table C.13 shows that the results on performance are qualitatively and quantitatively similar in this alternative specification where I assign them the payoff they would have received without the mistake.

Decision times provide additional support that performance differences are due to participants learning how much their partner likes them. I argue that behavior changes because the knowledge about interpersonal preferences changes collaboration. One alternative interpretation of the results is that participants become more interested in the characteristics of their partner when learning how they were ranked. So they increase attention to their partner's questionnaire responses and adjust their beliefs and behavior due to more careful inspection.[13] Following this argument, I would expect participants in *Info* to spend significantly more time inspecting their partner's profile before proceeding to the task than participants in *No-Info*. However, participants in *Info* spend a similar amount of time on the screen where they learn who their partner is (24.8 seconds) compared to *No-Info* (22.5 seconds). This makes it unlikely that results are driven simply by increased attention to their partner's characteristics. Therefore it seems to be the knowledge about the partner's preferences that changes how well teams with the same underlying preferences perform. While some of these revelations increase performance, others decrease performance, resulting in very similar average performance across treatments (see Appendix C.2.2 for a detailed analysis).

3.3.4 Underlying Channels

Ex-ante Beliefs and Preferences

I investigate whether participants' expectation about the upcoming task depend on how much they like their partner and how much their partner likes them in *Info*.[14] I separate beliefs about the efficacy of their upcoming interactions and preferences for working with their assigned partner.

First, participants expect their team to be more successful the more both team partners like each other. This is shown in Table 3.3. They especially expect to be significantly more successful with partners they like (Column 1). Together with the insignificant (but positive) point estimate on the preferences of their partner (Column 2), this translates into a significantly more positive evaluation of the team success when partners like each other. They also believe partners who like them to exert higher effort (Table C.4). In contrast, participants think that their own behavior will not be affected by either their own or their partner's preferences (Table C.2). In neither of these three assessments, participants consider the difference between their own and their partners' preferences to be meaningful.

Result 4. *Participants expect to be more more successful when being in a team where partners like each other. Whether or not their preferences align with their partner's preferences does not affect their expectations.*

[13] Holding such beliefs seems plausible, despite the fact that questionnaire responses are not correlated with actual performance as shown in Table C.14.

[14] In Appendix C.2.3, I show how these measures compare between *Info* and *No-Info* and corroborate the findings of this Section.

Table 3.3: Team Success [*Info*]

	Expected Team Performance			
	(1)	(2)	(3)	(4)
Preference for partner (1-4)	.620***			
	[.177,1.064]			
Partner's preference (1-4)		.364		
		[-.099,.827]		
Joint Liking (2-8)			.447***	.442***
			[.147,.747]	[.140,.744]
Dissimilarity Liking (0-3)				.092
				[-.451,.636]
Mean Dep. Var.	7.358	7.358	7.358	7.358
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
N	120	120	120	120

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is the amount bet on the team being in the top 25%, with monetary values recoded on a scale of 0-10. The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Second, participants do not expect to receive a higher task utility from performing the *Guesstimation* task in a team where partners like each other. They are equally willing to work alone in a payoff-irrelevant task, irrespective of how much they like their partner (Table 3.4, Column 1) or how much their partner likes them (Column 2). They are not more prone to prefer working alone when there is a larger difference between their own and their partner's interpersonal preferences (Column 4).

Result 5. *Participants expect to derive a similar utility from performing the Guesstimation task with someone whom they like or who likes them. Whether or not their preferences align with their partner's preferences does not affect their expectations.*

Table 3.4: Task Utility [*Info*]

	Preference for Working Alone			
	(1)	(2)	(3)	(4)
Preference for partner (1-4)	-.002			
	[-.351,.347]			
Partner's preference (1-4)		.057		
		[-.299,.413]		
Joint Liking (2-8)			.024	.014
			[-.212,.261]	[-.223,.252]
Dissimilarity Liking (0-3)				.195
				[-.233,.623]
Mean Dep. Var.	5.475	5.475	5.475	5.475
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
N	120	120	120	120

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable indicates the willingness to pay to perform another round of the task with a fixed payoff alone (as opposed to with their team partner). The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Communication

Effective communication in sharing information, discussing approaches, and collaborating is key to success in the *Guesstimation* task. Because participants face limited time to solve the problem, there is a clear trade-off between working on parts of the task individually and communicating with their partner. I find strong evidence that extensive communication is not necessary for success. The correlation between the quantity of communication, measured by the total number of characters, and team performance is indeed negative in *No-Info* where preference revelation cannot affect communication patterns ($r = -0.3438$; Pearson's correlation coefficient). Successful teams typically require less than 1000 characters to reach a solution. In contrast, less successful teams exchange up to 2000 characters. While this should not be taken as evidence that less communication is conducive to performance, it does show that coordination on a common solution can be achieved with little communication.[15]

[15] Appendix Table C.12 shows in Column (1) that this also holds for the ex-post evaluation of the collaboration. How much teams communicate is not related to how positive participants evaluate the collaboration experience.

Changes in collaboration behavior are the only channel through which the revelation of each other's preferences can have a causal effect on performance. Since collaboration only happens via chat, any changes *should* be observable in the (written) communication. Differences in communication patterns may be due to changes in effort, shifting roles within the team, or changes in willingness to challenge the partner or compromise on a solution. Whatever the underlying reason for the differences in performance, it must operate through joint communication.

I find suggestive evidence that changes in communication patterns through the revelation of preference information drive the performance differences. Columns (1)-(3) of Table 3.5 consistently show that those who like each other communicate more in *Info*. Columns (1) & (2) also reveal that teams with *dissimilar liking* communicate less in *Info*, although the effect is imprecisely estimated, which is also the case when analyzing communication patterns in isolation in *Info* (Table C.8. Table C.9 shows no effects of the revelation of preferences on either communication asymmetry, turn-taking, or the time until the first message is sent.

Table 3.5: Team Communication [across Treatments]

	Total Length of Chat Messages		
	(1)	(2)	(3)
Joint Liking (2-8)	-38.819*	-39.233*	-42.086*
	[-84.267,6.628]	[-84.141,5.675]	[-87.848,3.677]
Joint Liking X Info	56.682*	61.796*	65.090*
	[-10.782,124.146]	[-5.048,128.639]	[-3.954,134.134]
Info	-350.430**	-259.216	-296.205
	[-696.427,-4.434]	[-617.546,99.114]	[-674.038,81.628]
Dissimilarity Liking (0-3)		-3.019	-5.015
		[-105.351,99.313]	[-110.085,100.054]
Dissimilarity Liking X Info		-94.906	-90.746
		[-230.384,40.572]	[-231.813,50.321]
Mean Dep. Var.	943.72	943.72	943.72
Team Controls	Yes	Yes	Yes
Performance Controls	Yes	Yes	Yes
Problem FE	No	No	Yes
N	120	120	120

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The values in square brackets represent the 95% confidence intervals. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Joint Liking X Info* interacts this variable with a treatment indicator for *Info*. *Dissimilarity Liking* is the absolute difference between both partners' preferences, *Dissimilarity Liking X Info* captures the interaction with the treatment *Info*. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants face.

Result 6. *Teams where partners know they like each other communicate slightly more, those who know about their misaligned preferences communicate slightly less.*

Ex-post Evaluation

Participants do not evaluate their interaction ex-post differently depending on the interpersonal preferences of both team partners. Whether team partners liked each other or whether these preferences were misaligned does not determine how they judge the interaction retrospectively. The same is true in terms of their decision on whether to continue working with their team partner. In Table C.10, I show that neither psychological safety, the group climate, task, enjoyment, the willingness to compromise, nor the fear of disappointing others are affected by both partners' preferences.[16] Participants are also equally inclined to perform another identical round of a *Guesstimation* task with their partner, irrespective of their preferences. In *Info*, 60 percent want to continue working with their team partner. Table C.11 shows that the preference for working alone remains unaffected by whether team partners like each other or not, and whether their preferences are aligned or not. Taken together, these results provide suggestive evidence that participants did not experience greater utility from working with a partner who liked them, nor do they believe they were more successful with such a partner. In that sense, participants learned through the collaboration that their ex-ante expectations of higher team performance when liking each other were not met. At the same time, participants in teams with dissimilar preferences did not seem to experience collaboration more positively either, nor did they believe to be more successful and hence to continue with their partner.

Result 7. *Participants evaluate the collaboration in the Guesstimation task similarly, irrespective of whether they were in a team where partners liked each other, disliked each other, or had misaligned preferences.*

3.4 Discussion

In this study, I analyze how interpersonal preferences affect team performance in complex problem solving. To do so, I rely on a non-routine task that captures 21st century

[16]Columns (2)-(5) in Table C.12 show consistency in the responses between the different sets of questions, providing evidence that the null results are not driven by a lack of attention to these (non-incentivized) questions.

skills and is used frequently to assess the quality of job candidates. This *Guesstimation* task allows me to create a team task in which complementary knowledge and skills can come into play when participants communicate with each other effectively. Hence, I analyze whether interpersonal preferences change the effectiveness of teams in a task with high external validity.

I find that interpersonal preferences matter for effective teamwork. When individuals know each other's preferences, dissimilar preferences are conducive to increased team performance. That is, teams where one partner likes the other more perform better. However, teams where partners like each other do not perform better than teams where they do not. Team members do not anticipate this pattern. Before conducting the task, those in a team where partners like each other expect to be more successful in the task. I provide suggestive evidence that more effective communication (as opposed to more communication) drives the performance results and establish that the results are not due to homophily and sorting.

Relying on an experimental setting allows me to infer participants' preferences and manipulate information structures. Incentivizing truthful preference submission and tightly controlling information structures would be impossible in the field. This comes at a cost. First, I elicit preferences that are meaningful as they shift behavior (see also Opitz & Schwaiger, 2023a), but are not as strong as they can be in real-word settings. A preference formed on the basis of five answers to questions is likely to be less strong and more malleable than a preference formed on the basis of hearing the opinions and observing the behavior of others. Second, the *Guesstimation* task abstracts from other dimensions where interpersonal preferences may matter in the work context. I focus on the performance aspect during the task, while working with friends may be especially important when it comes to fair attribution of responsibility, success, and failure after task completion (e.g., Jin et al., 2019).

This study extends the findings of Opitz and Schwaiger (2023a), which shows the relevance of being liked for cooperation. In line with Opitz and Schwaiger (2023a), I find that participants expect higher effort from a partner who likes them, and believe that this will translate into higher performance. This is the case even though I am studying a collaborative environment with aligned interests rather than a cooperative environment with conflicting individual interests, and I conceptualize interpersonal preferences slightly differently. In Opitz and Schwaiger (2023a), preferences for team partners are based on the desirability to perform the cooperative game with them. In contrast, this study attempts to capture preferences that are not tied to the specific task

–instead they refer to a payoff-irrelevant interaction. In that sense, this study is closer to a definition of *liking* than to *task-specific partner preferences*.[17] Despite the more favorable ex-ante beliefs, objective performance is not better in teams where partners like each other. While in stylized one-shot interactions, these beliefs closely map into actions (and payoffs), the determinants of success in the collaborative problem solving environment of this study are more complex. This highlights the importance of better understanding the production function in such complex tasks and analyzing communication as a key component.

Overall, these findings raise important questions on how to make communication effective. Charness, Cooper, and Grossman (2020) shows that imposing communication costs can improve team performance on logical problems through decreasing the quantity and improving the quality of messages, Girotra, Terwiesch, and Ulrich (2010) shows the benefits from limiting communication temporarily. I also show that the quantity of communication does not necessarily translate into higher success. In addition, I find that the quantity of communication is largely unrelated to the self-reported collaboration experience of participants. While Girotra, Terwiesch, and Ulrich (2010) and Charness, Cooper, and Grossman (2020) change the nature of communication exogenously, my experimental condition holds everything constant except for the piece of information on how one was ranked by the team partner. This, in turn, influences how individuals communicate to combine their knowledge and solve the problem. Interestingly, participants did not report a better or worse collaboration experience depending on both partners' preferences.

The results suggest several important organizational implications. Self-selected teams would most likely have been detrimental for performance if preferences were either known or if individuals learned each other's preferences in the process of forming teams. Because teams are formed randomly in the experiment, I do not observe performance differences between the two information structures on average. Still, if participants had been allowed to form teams on their own, there had been more teams in which both partners liked each other –and these teams turned out to be less successful in the *Info* treatment. This has two underlying mechanism. First, the assignment would not be random anymore, take the preferences of both partners into account,

[17] In this sense, the study relates more to papers that investigate the effects of friends at the workplace, including Bandiera, Barankay, and Rasul (2005) and A. Ashraf (2022).

and lead to higher fraction of individuals who work with their preferred partner.[18] Second, preferences themselves are likely to change because people take their partner's preference into account when forming a team, since they expect a partner who likes them to exert more effort. This implies that having teams that self-select based on how much team members like each other could be detrimental to overall performance without teams expecting this when they form. Teams where one or both partners know they are not necessarily each other's favorites perform better. In this sense, these findings contribute to a better understanding of how teams should be formed, and which preferences –if known– should be revealed.

[18]For example, a counterfactual assignment of participants to teams via the Deferred-Acceptance Algorithm (Gale & Shapley, 1962) instead of a random assignment would have increased the fraction of individuals who perform the *Guesstimation* task with their most preferred partner from 23 percent to 60 percent.

4

Gendered Access to Entrepreneurial Finance

The Role of Team Formation, Idea Quality, and Business Implementation

4.1 Introduction

Access to finance plays a crucial role in unleashing the potential of entrepreneurial ideas and transforming start-ups into successful businesses. While many firms face financing constraints (Carpenter & Petersen, 2002; Banerjee & Duflo, 2014), these constraints are more pronounced for female entrepreneurs: Women are less likely to have the necessary financing to start a business (OECD, 2017), they face challenges in attracting external equity (Guzman & Kacperczyk, 2019; Ewens & Townsend, 2020;

*This chapter is based on joint work with Vojtěch Bartoš, Silvia Castro, and Kristina Czura.

Hebert, 2020), and their constraints to debt finance are more pronounced. Female entrepreneurs secure smaller loan amounts (Bellucci, Borisov, & Zazzaro, 2010; Agier & Szafarz, 2013; Demirguc-Kunt et al., 2018; Bartos et al., 2023), pay higher interest rates (Asiedu, Freeman, & Nti-Addae, 2012), are more likely to be denied a loan (Morazzoni & Sy, 2022), and are required to provide more loan guarantees (Brock & De Haas, 2023). Designing targeted policies to close the observed gender gap requires an understanding of whether the gap is driven by demand or supply factors, and to identify potential existing gender bias and its underlying mechanisms.

In this study, we focus on gender bias and its underlying mechanisms on the supply side of access to finance. Specifically, we analyze whether loan officers' assessment of a start-up's performance depends on the gender of the entrepreneurs and the size of the team. Using a lab-in-the-field experiment in Uganda, 451 loan officers of a large bank for entrepreneurial finance evaluate real-life business pitch decks, a short presentation of a business idea, where the gender and the team composition has been randomly manipulated by us. This design enables us to complement Brock and De Haas (2023) in isolating supply-side gender bias, and to disentangle underlying mechanisms for a gender bias in loan officers' evaluations. We distinguish gender differences in the evaluation of the business idea itself from the evaluation of the implementation challenges and capabilities of the entrepreneur.

Our setting is very suitable to answer this question. First, we study debt financing, the predominant source of finance for start-ups in low-income countries where venture capital is scarce (AVCA, 2022; Jaoui, Amoussou, & Kemeze, 2022).[1] Second, loan officers are a relevant sample since they are the first point of contact of an entrepreneur with the bank. Their role is to assess the creditworthiness of applicants based on often incomplete information about the business and the entrepreneur. However, predicting entrepreneurial success is a difficult task (Fafchamps & Woodruff, 2017; McKenzie & Sansone, 2019); and the resulting, partially subjective judgement leaves ample room for bias. Third, the controlled but natural RCT-like design allows us to identify causal effects of entrepreneurs' gender on loan officer decisions. These decisions are not confounded by other characteristics of the business or the entrepreneur that may otherwise affect decisions without being observable to the researcher. In particular, our design allows us to control for any factor arising from the demand side for access to

[1] In 2022, the value of venture capital deals in Africa was 5.2 billion USD. While this constitutes a three fold increase over the previous year, the total amount is only around one percent of the total global value of venture capital deals (AVCA, 2022). Moreover, most of the venture capital deals are concentrated in countries such as South Africa, Egypt, Kenya, and Nigeria (Economist, 2022).

finance, so we can focus exclusively on the supply side.[2]

Our design establishes the causal role of entrepreneurial gender and team formation on loan officers' business evaluations. We present all business pitches with a founder, who has developed the business idea, and an implementer, who executes the business idea. This serves two purposes. First, we disentangle whether any differential evaluation of the business potential stems either from gender bias in the assessment of the business idea, or from the perceived entrepreneurial ability and constraints an entrepreneur faces when operating a business. To do so, we experimentally manipulate the gender of the founder and the implementer. Second, we compare gender biases in evaluations of individual entrepreneurs (i.e., sole proprietorship) versus teams of two entrepreneurs by varying whether the founder and the implementer are the same or two different entrepreneurs. While the first part is important for improving our understanding of gender bias in access to finance, the second part is crucial to understanding the role of forming a team for entrepreneurial success, in particular in low income countries (Hsieh & Olken, 2014; Ulyssea, 2018). Most enterprises in these settings hardly grow and often remain one-person businesses (Calderón, Iacovone, & Juarez, 2017). Gender bias in access to finance for such businesses may contribute to the under-representation of female businesses among larger firms in less developed countries. At the same time, entrepreneurship education programs, as well as incubators and accelerators that support startups, strongly emphasize the formation of entrepreneurial teams for sustained business success.[3] A potential gender difference in the returns to entrepreneurial team size would suggest that these approaches address one key barrier to the growth of female businesses.

Loan officers evaluate the business pitched along several dimensions. First, loan officers can invest into each presented business pitch and their return depends on business survival. Second, they select the best performing business among the evaluated pitch decks. Third, loan officers can engage in additional costly screening to assess a business. In line with statistical discrimination (Phelps, 1972; Arrow, 1973; Aigner & Cain, 1977), if loan officers are missing information about businesses from a certain

[2]Demand side factors such as differences in risk attitude, or willingness to ask or negotiate may also be an explanatory factor for observed gender differences in access to finance (e.g., Bowles, Babcock, & Lai, 2007; Croson & Gneezy, 2009; Card, Cardoso, & Kline, 2015; Niederle, 2016).

[3]From early entrepreneurship education programs to accelerators, the formation of an entrepreneurial team is often required (e.g., the "LaunchX" program for high school students or the "Berlin Startup School Accelerator"). Even the world's largest and most successful incubators, such as "YCombinator" or "Techstars", strongly encourage the formation of an entrepreneurial team prior to application. The same is true for successful African incubators like "The Baobab Network".

class of entrepreneurs, they should be willing to invest resources into obtaining this information as long as the information is expected to change the prior belief. The evaluations above are incentivized based on the real-life business performance of the start-ups around two years after pitching the business idea in a business plan competition. Additionally, we elicit non-incentivized probabilistic beliefs about business success and a subjective measure of the perceived quality of the business idea. The outcome measures are informative about the entire distribution of loan officers' evaluations of the future business profitability. This allows us to relate our findings to both equity and debt financing, which are arguably more concerned with different aspects of firm performance: business success for high returns on equity and firm survival for loan repayment.

We find a sizable gender bias for businesses of individual entrepreneurs. Loan officers invest around 8 percent less in businesses by female entrepreneurs, they are 7 percent less likely to select a pitch as the best businesses when it is pitched by a female entrepreneur, and they consider the probability of failure to be 17 percent higher when the same idea is pitched by a female entrepreneur. These effects are more pronounced among loan officers who hold gender biased attitudes, who are less experienced, and who are female. The observed premium for individual male entrepreneurs over individual female entrepreneurs is in line with the type of clients loan officers usually interact with: 70 percent of them are male. Examining the remaining outcomes allows us to examine potential channels driving the result. First, no gender difference in the subjective evaluation of business idea quality implies that the gender bias is not driven by animus against ideas developed by women. Second, the absence of gender disparities in the costly screening process suggests that loan officers do not perceive they lack information about female-led businesses. Instead, we conjecture that the observed gender bias stems from differential assessments of women's entrepreneurial ability or external constraints when running a business.

In contrast, we do not observe a similar gender bias in the evaluation of teams of two entrepreneurs. Loan officers do not invest differently in businesses when they were founded or implemented by a female member of the entrepreneurial team. This null result is not caused by lack of variation, low power, or lack of attention and effort. Further, investment behavior is correlated with other proxies of business idea quality at the individual level, so we are confident in the measure's validity. We do find some indication of a different type of a bias at the top, a penalty in the evaluation of mixed-gender teams. A business pitch from mixed-gender teams is less likely to be selected as the best performing business, although the result does not always reach statistical significance. Comparing teams to individual entrepreneurs shows that loan officers do not evaluate their business pitches differently, on average. Our results allow us to rule out that the contrasting results for individuals and teams are driven by relative unfamiliarity or by different preferences or beliefs of loan officers about either type of a business. For one, loan officers do not request more information about teams. And second, loan officers evaluate the profitability of teams and individuals equally. Taken together, despite the fact that almost all applications the bank typically processes are from individual applicants, teams would not likely suffer any penalty, nor receive a premium.

We contribute to three strands of the literature. First, we contribute to the literature on gender discrimination in entrepreneurial finance. Previous work has documented an investor bias against female entrepreneurs (Guzman & Kacperczyk, 2019; Ewens & Townsend, 2020; Hebert, 2020) and that female borrowers face tighter credit availability or less favorable loan terms (Muravyev, Talavera, & Schäfer, 2009; Bellucci, Borisov, & Zazzaro, 2010; Asiedu, Freeman, & Nti-Addae, 2012; Agier & Szafarz, 2013; Alesina, Lotti, & Mistrulli, 2013; Mascia & Rossi, 2017; Morazzoni & Sy, 2022). Recent experimental work has pinpointed to loan officers' gender bias as a source of gender disparities in entrepreneurial finance (Alibhai et al., 2019; Brock & De Haas, 2023; Zhang, 2023). Closest to our study, Brock and De Haas (2023) provide causal evidence for gender discrimination in entrepreneurial lending. Using data from a lab-in-the-field experiment in Turkey, they document that loan officers indirectly discriminate against female loan applicants by requesting more loan guarantees. These effects are concentrated among female businesses in traditionally male industries, suggesting that gender stereotypes drive this discrimination. Yet, they do not find direct discrimination against female applicants. We contribute to this literature by cleanly identifying supply side factors for gender bias. Further, we advance this literature and investigate the mechanisms underlying the potential bias beyond differential treatment of male and female loan seekers. First, we examine whether differential treatment is the result of discrimination when evaluating business ideas or of different beliefs about women's abilities and constraints in implementing the business idea. Second, we study differences in screening efforts for male and female entrepreneurs. This differentiation is particularly important for tailoring policies to increase women's participation in credit markets. Finally, we examine how team size interacts with gender bias.

Second, we contribute to the literature on the determinants and biases in predicting business success and how this affects access to finance. The prediction of entrepreneurial success is a difficult task for both human experts and state-of-the-art machine learning approaches (Fafchamps & Woodruff, 2017; McKenzie & Sansone, 2019). Yet, loan officers' ability to properly evaluate potential business success is key for viable entrepreneurial finance, in particular when information on the business and loan applicant is scarce. Information scarcity is prevalent in many low-income countries without existing credit registries or systematic business accounts (Djankov, McLiesh, & Shleifer, 2007). Subjective evaluations are prone to gender biases (M. Lee & Huang, 2018) and information-scarce credit markets allow for animus driven behavior and favoritism (Blanchflower, Levine, & Zimmerman, 2003; Younkin & Kuppuswamy, 2018). Cole, Kanz, and Klapper (2015) show that high-powered incentives induce loan officers to provide more effort into screening loan applications, while volume-based incentives can lead loan officers to overlook valuable soft information (Agarwal & Ben-David, 2018). Even status symbols like obesity (which is perceived as a reliable signal of wealth in many low-income settings) are affecting loan approval decisions in such a low-information setting (Macchi, 2023). Reducing information frictions between the borrower and the loan officer by cultural proximity (Fisman, Paravisini, & Vig, 2017) or the same gender (Beck, Behr, & Madestam, 2018) improves access to finance, loan conditions, and repayment.[4] We contribute to this literature by introducing gender bias as a possible confounding factor in the evaluation of business potential. Unlike Beck, Behr, and Madestam (2018), who document homophily and in-group favoritism in the loan terms for first-term borrowers in an Albanian bank, we do not observe preferential in-group treatment in business evaluation. One central difference in our study is that we exclude any effects arising from the demand side, which suggests that in-group favoritism emerges through direct communication during the screening process.

Last, we also extend the understanding of underlying sources of gender bias in access to finance. Typically, studies aim at providing evidence supportive of either taste-based (Becker, 1957) or statistical-based (Phelps, 1972; Arrow, 1973; Aigner & Cain, 1977) types of discrimination (Gonzales Martinez et al., 2020; A. M. Montoya et al., 2020; Macchi, 2023). Our subjective assessment about the business idea and the incentivized information acquisition task allows us to make inference about the role of taste-based and statistical-based discrimination. On top, our novel design allows us to separate loan officers' evaluations of the idea quality from the business implementa-

[4] Similarly, in the private equity context, similarity between the members of the start-up team and the venture capitalist has been shown to increase access to finance (Franke et al., 2006).

tion capacity. In doing so, we locate the source of gender bias in the perceptions of the implementation ability of female entrepreneurs, rather than in the perceived quality of their ideas. Understanding these underpinnings of the bias allows policymakers to design effective tools to address the inequalities.

4.2 Experimental Design

In our lab-in-the-field experiment in Uganda, loan officers evaluate a set of business pitch decks from start-ups. Our objective is to examine whether these evaluations differ along two dimensions: the gender of the entrepreneur and the formation of entrepreneurial teams. With regard to gender, our study design enables us to differentiate whether any observed differences stem from varying assessments of the business idea itself or of the implementation challenges and capabilities of the entrepreneur. Business ideas are evaluated using two measures: first, determining whether to invest in the showcased business and second, the selection of the most promising business among all the presented pitch decks. These decisions are incentivized based on the real-life performance of the start-ups from the business pitch. Additionally, we analyze differences in the effort to screen the start-up businesses and non-incentivized beliefs about the quality of the business idea and the business success.

4.2.1 Sample and Setting

We partner with a large Ugandan commercial bank that specializes in lending to small-scale businesses and entrepreneurs. We selected 28 branches with more than eight loan officers that are feasible to reach in a one-day trip from the capital, Kampala, or other major Ugandan cities. Our sample covers 35 percent of all bank branches and about 45 percent of all bank loan officers. In each branch, we invite all loan officers who handle business related loans for participation in our experiment without informing them about the content of the study. Participation in the study is voluntary. Our final sample consists of 451 loan officers. Loan officers are on average 34 years old, 55 percent are female, and they have an average of 6.7 years of experience in the position.

The business pitches the loan officers evaluate in our experimental sessions have been presented by graduates of an entrepreneurship academy at a business plan competition. In a related study, we evaluate the impact of entrepreneurship academies on the business performance of these start-ups (see Bartos et al., 2023). Entrepreneurship academies are run at several Ugandan universities with university students interested in pursuing entrepreneurial careers. We follow these nascent entrepreneurs from their application to the entrepreneurship academy until around two years after they have completed the training and participated in the business plan competition. For each business pitch, we have detailed information on the team of entrepreneurs and their business performance two years after the business idea was pitched. The evaluation decisions of the loan officers in our experiment are incentivized based on the real-life information on business performance from the pitching start-ups.[5]

We selected five pitch decks from the sample of 58 pitch decks that were pitched at the business plan competition. First, we excluded pitch decks that did not provide enough information about (expected) business performance for loan officers to make an informed decision. Second, we excluded ideas that were clearly perceived as either male or female businesses in our pre-testing. We additionally validated our identifying assumption that participants cannot infer the gender of the entrepreneurial teams solely by looking at their idea in a survey with 38 Ugandan university students.[6] We do not detect strong beliefs about the gender of the business owner(s): While actual pitch decks by males or male teams were evaluated as more male, relative to female or female team businesses ($p = 0.064$), the modal belief is that the idea came from a team with an equal proportion of men and women (63 percent).

4.2.2 Conceptual Framework

For our evaluation experiment, we standardize the presentation of the business pitches. All pitch decks are presented with a founder and an implementer (i.e., the *CEO* or *manager*). We make it clear that the founder has developed the business idea, while the implementer executes the business idea. The founder and the implementer may or may not be the same person.

We model the perceived business performance (B) as a function of both the quality of

[5] We have received informed consent from all founders that their pitch decks can be used for the purposes of a research study.

[6] To do so, we first removed all identifiers of the actual entrepreneurs from all pitch decks. Then, each student evaluated a randomly selected subset of 20 pitch decks out of the full sample of 58 pitches. We asked the students whether they thought "the owner or the group of owners is more likely to be [all male / mostly male / male and female in equal proportion / mostly female / all female]".

the business idea (Q), as well as the implementation of the idea (I). To understand gender-specific business evaluations, both parameters are gender-specific $g = \{M, F\}$, such that we have $B(Q_g, I_g)$. Varying the gender of both dimensions allows us to disentangle whether a gender-specific business evaluation originates from a differential evaluation of the idea quality itself, or from different perceptions about the potential of an entrepreneur to implement it. While differences in idea quality indicate gender bias in the evaluation of the business idea, different beliefs about the potential to implement an idea may either stem from gender-specific beliefs about (external) constraints or the (personal) ability to implement the idea. To understand how team formation influences business evaluations, in particular if there are gender-specific evaluations for female and male businesses, we vary whether the businesses are founded by an individual entrepreneur or by a team of two entrepreneurs.

4.2.3 Gender and Entrepreneur(s) - Exogenous Variations

We exogenously vary two components in the evaluation of the business performance: First, we vary the gender of both the founder and the implementer and compare the loan officers' business evaluations across these four founder-implementer gender combinations. Second, for founder-implementer combinations with the same gender, we vary whether the business is proposed by a team of two entrepreneurs or an individual entrepreneur.

Specifically, a loan officer i evaluates the business success $B_i^p(Q_g, I_g)$ of pitch deck p. Every loan officer sees the same five pitch decks in the same sequence. We randomly assign the founder-implementer gender combinations for each pitch deck across loan officers resulting in a between-subject design. This means that for a pitch deck p, a loan officer i either evaluates $B_i^p(Q_M, I_M)$, $B_i^p(Q_M, I_F)$, $B_i^p(Q_F, I_M)$, or $B_i^p(Q_F, I_F)$.

We vary the gender of the founder and implementer in the following way: We remove all personal information from each pitch deck, i.e., all information on the entrepreneurs proposing this business pitch. In the next step, we create four versions of the pitch deck. We assign a founder-implementer gender combination to each of the four anonymized pitch deck clones. For this, we vary the dimensions of the *founder's* gender and *implementer's* gender (male vs. female) in a 2x2 design. The gender of the founder and implementer are revealed by their names on the pitch deck (without photos or any additional information). We made sure that the used names are clearly associated with one gender only and that ethnicity, socio-economic status, or other characteristics could not be inferred from them.[7]

We then vary the team formation of entrepreneurs by introducing a fifth pitch deck clone in which the founder and the implementer are the same person, i.e., an individual entrepreneur. With a probability of 50 percent, the entrepreneur is male, and with 50 percent female. For each pitch deck p, we randomly assign the five-pitch deck clones across participants. We assign four-pitch deck clones of teams of two entrepreneurs with four different founder-implementer gender combinations and one pitch deck clone of an individual entrepreneur with a random gender distribution.[8]

4.2.4 Evaluation of Business Ideas

We elicit both incentivized and unincentivized decisions to evaluate the business ideas. Our two main outcome variables are the incentivized measures *Investment decision* and *Best business*. Both decisions are incentivized based on the real-life business performance of the pitching start-up as follows. *Investment decision* is a continuous variable stating the amount that participants invest in each pitch deck. Loan officers are endowed with 5,000 UGX that they can invest into each business (in increments of 500 UGX).[9] The investment amount is doubled if the corresponding real-life business reports positive profits two years after pitching the idea. The investment is lost if this business reports negative profits or does not exist anymore. Investors keep the non-invested part of the endowment. This outcome captures loan officers' ability to predict business survival for each pitch deck.

After all pitch decks have been evaluated individually, participants select the *Best business* which they believe has generated the highest profits. Loan officers receive a fixed 5,000 UGX bonus payment if they identify the real-life business with the highest profits and nothing otherwise. This outcome captures loan officers' ability to identify the

[7]We selected the names we assigned to the pitch decks as follows. The name is either a real name of a team member or a name of another participant of the entrepreneurship academy. We tested a set of 30 names of academy graduates among 10 Ugandan natives on whether ethnicity, religion, socio-economic status, or other characteristics could be inferred from these names. We selected five sets of names (two female names and two male names each) that are general enough such that respondents could not infer anything about ethnicity, socio-economic status, or level of education. We excluded all names where gender was not clear to all respondents. All names we used are associated with Christian religion, so they are not confounded by religious identity either. See the list of all names is in Appendix Table D.1.

[8]Loan officer characteristics are balanced across the two gender realizations (see Appendix Table D.2). All other manipulations are within-subject.

[9]5,000 UGX correspond to around 1.28 EUR in December 2022 when the experimental sessions were implemented with an exchange rate of 3.858 UGX/EUR.

most successful business in terms of profits.

To understand whether potential gender bias is based on information scarcity or higher uncertainty about women's businesses, we provide loan officers with additional screening options. The investment payout is based on one selected pitch deck and loan officers have the option to purchase further information on the entrepreneurs and/or the business. Subsequently, they may revise their initial investment decision for the payoff-relevant pitch deck. We present a list of information items about the entrepreneur's background and the business.[10] Loan officers state which information they would need to best assess the pitch deck and they decide whether and which pieces of information they want to purchase. Each piece of information costs a fixed amount of 200 UGX.[11] The decision to purchase information is incentive compatible. Participants know that they will have a chance to revise their investment decision after the possibility to obtain the additional information. It is thus in their best interests to select the information they deem relevant. We generate the following outcome variables: (i) an indicator variable on whether a loan officer purchased any information, (ii) the number of information pieces purchased, and (iii) indicator variables on whether a loan officer purchased the information for each piece of information.

The following unincentivized measures are our secondary outcome variables. *Idea quality* is based on two survey questions for each pitch deck, i.e., whether the loan officer agrees that *the business meets a need or solves a problem in Uganda* and *that there is a market for this business idea in Uganda*. *Idea quality* is the average of both questions, which are answered on a scale from 0 to 100. *Beliefs about success* for each pitch deck are measured by the probability distribution across the three options that this business idea will either (i) fail within the first year, (ii) survive in the first year and make small profits, or (iii) survive in the first year and make large profits.

Finally, we asked three questions on gender norms (Scholz et al., 2014), and collected basic demographics of loan officers (gender, age, years of experience).[12]

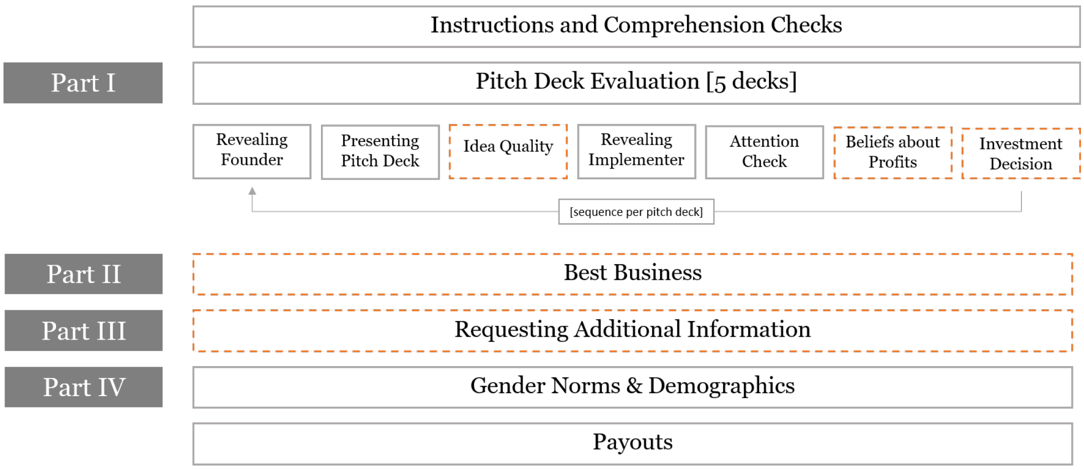
[10]The following items are displayed on the list: (1) All team members owning the business, (2) Professional references for business owners, (3) Professional experience of the business owners, (4) Professional network of the creators, (5) Financial support from family members this business has received, (6) External financing obtained, and (7) Volume of sales, revenues, and profit margins.

[11]As we only have a subset of this information available for the payoff-relevant pitch deck, loan officers only pay for those selected categories for which the information is actually available.

[12]We asked Questions R1c, R1d, R2b from Scholz et al. (2014), asking how much the respondent agrees with the following statements on a 5-point Likert scale coded from 1. ...completely disagree to 5. ...completely agree: (i) *A man's job is to earn money; a woman's job is to look after the home and family.*, (ii) *A job is all right, but what most women really want is a home and children.* , and (iii) *Family life suffers when the woman has a full-time job*.

4.2.5 Procedures

The experimental sessions were conducted at the branch offices of our partner institution. We visited the branches outside the regular business hours to avoid client traffic or other distractions during the experimental sessions. Experimental instructions were explained by the research team using flip chart illustrations (see Appendix D.1). We ensured loan officers' understanding of the instructions by two comprehension questions.[13] All decisions were collected in a survey on a digital device. We either used the loan officers' personal workstations or provided tablets for them. Loan officers could proceed through the survey at their own pace. A research team member assisted them if needed. The experiment was programmed in Qualtrics.



Notes. This figure presents an overview of the study design. The outcome measures are displayed in dashed boxes.

Figure 4.1: Design Overview

The experiment is organized in four parts (see Figure 4.1). In Part I, loan officers evaluate five pitch decks. Each pitch deck is presented in the same sequence: loan officers 1) learn who the founder of the business idea is, 2) see the business pitch, 3) evaluate the idea quality qualitatively (note that this is independent of the identity of the implementer), 4) learn who the implementer of the idea is, 5) pass an attention check on the gender allocation for the founder and implementer[14], 6) indicate their

[13]Questions are: (i) "Imagine that you invest 2,500 UGX to the business and keep 2,500 UGX. The business reported that it still exists and makes profits. How much do you have in total?", and (ii) "Imagine that you invest 4,000 UGX to the business and keep 1,000 UGX. The business reported that it does not exist anymore. How much will you have in total?". Loan officers were only allowed to proceed when answering both questions correctly. If they failed to answer the comprehension questions correctly, the instructions were repeated.

[14]Participants are presented with three statements about the description of the idea and the founding team. Only after indicating the correct answer, participants can progress.

probabilistic beliefs about business success, and 7) state the amount they would like to invest in the business. All pitch decks are presented in the same sequence to not confound the gender variation within each pitch deck with potential order effects from variations in the sequence. In Part II, loan officers select the business pitch they think has performed best (i.e., generates the highest profits). In Part III, loan officers can request additional information for the selected —payoff-relevant— pitch deck after learning about the surprise option to revise their investment choice for that pitch deck. Lastly, in Part IV, we elicit gender norms and socio-demographics. See the full wording of the survey and screenshots in Appendix D.2.

Loans officers' final payoff consists of four components. First, loan officers receive a participation fee of 5,000 UGX. Second, loan officers receive an initial endowment of 5,000 UGX; they keep the part that they did not invest in the payoff-relevant business, and the return from the amount invested into this business. Third, the amount for the purchased information is subtracted from the participation fee. Fourth, loan officers receive a bonus payment of 5,000 UGX for correctly identifying the best performing business. Earnings were delivered via mobile money soon after the experimental session was finished. Average payouts for the one-hour session were 13,472 UGX, which is in the range of the average hourly wage (about 10,000 UGX for the loan officers in our sample).

4.3 Results

We document a bias against individual female businesses as opposed to individual male businesses, no gender bias for teams of entrepreneurs, and suggestive evidence of a bias against mixed-gender teams. The bias against individual female businesses seem to be driven by beliefs about implementation ability or implementation constraints, possibly magnified by individually held gender stereotypes. We first present results for individual entrepreneurs. Second, we turn to results for entrepreneurial teams. Finally, we compare decisions for individual entrepreneurs and teams.

4.3.1 Gender Bias in the Evaluation of Individual Entrepreneurs

Loan officers exhibit bias against individual female entrepreneurs in their investment decisions. We report results in a regression specification in Panel A of Table 4.1. Each loan officer evaluates one pitch deck of an individual entrepreneur, half of the these are female, the other half male entrepreneurs. The dependent variable is the amount (in UGX) invested in each pitch deck. The regressions report coefficients for indicators for a pitch deck of a female entrepreneur. The excluded category is a pitch deck with a male entrepreneur. Each regression includes pitch deck fixed effects. We report robust standard errors. On average, loan officers allocate 245 UGX less to female businesses ($p = 0.040$). The effect represents seven percent of the average amount invested in business ideas of male entrepreneurs (3,490 UGX). The result is robust to including individual controls (Table D.3) and to restricting the sample to the more attentive loan officers as proxied by number of clicks required to answer comprehension questions correctly (Table D.4).[15] The gender difference in investment rates is present throughout the entire choice distribution (Figure D.1). Our results imply that loan officers expect gender differences in the ability of businesses to generate profit.

Table 4.1: Investment and Best Business Decision [Individual Entrepreneurs]

Panel A: Investment		Gender bias		LO gender		Experience	
		Low	High	Female	Male	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Entrepreneur	-245.24**	-139.47	-355.85**	-289.02	-202.74	-265.18	-233.11
	(119.28)	(169.03)	(170.25)	(179.75)	(160.33)	(184.96)	(161.94)
Mean Dep. Var.	3,490.7	3,342.1	3,656.9	3,333.3	3,609.8	3,442.7	3,529.2
Observations	451	234	217	201	250	199	252
Panel B: Best Business							
Female Entrepreneur	-0.07*	-0.03	-0.11**	-0.10*	-0.04	-0.13**	-0.04
	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
Mean Dep. Var.	0.26	0.25	0.21	0.24	0.28	0.33	0.21
Observations	451	234	217	201	250	199	252

Notes. OLS Regressions of the dependent variable *Investment* (Panel A) or *Best Business* (Panel B) on the gender of the individual entrepreneur who founded and implemented the business. Panel A reports the incentivized decision of how much to invest in the pitch deck business from 0-5,000 UGX. Panel B reports the incentivized decision of which of all the businesses is the best idea of all the ones seen; it is a probability. Column (1) reports the average effect, and Columns (2)-(7) split the observations according to different relevant observable characteristics. Columns (2)-(3) split by gender bias following International Social Survey Programme gender bias metrics. It incorporates three questions: *A man's job is to earn money; a woman's job is to look after the home and family, a job is alright, but what most women really want is home and children,* and *family life suffers when the woman has a full-time job*. The sample is split at the median. (4)-(5) are split according to the self-reported gender of the respondent, and (6)-(7) are split according to the median experience level. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust and reported in parentheses. The table includes pitch deck FE in both Panels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[15] We remove the choices of ten percent of individuals who clicked most on the survey page when answering comprehension questions.

In line with the gender bias observed across the investment distribution, we also observe gender bias in the probability of female businesses being selected as the best performing business. Using a linear probability model with a similar regression framework, a female entrepreneur's business is 7 percentage points ($p = 0.052$) less likely selected as the best relative to an otherwise identical business with a male entrepreneur (Table 4.1, Panel B), corresponding to a reduction of 30.1 percent of the average of a male entrepreneur's business (23.3 percent). This effect is robust to regression specifications with individual controls (Table D.3). When focusing on the sample of attentive participants, we lack statistical power to detect significant effects, but the point estimate is very close to that detected in the full sample and they cannot be distinguished statistically (Table D.4). These results are also reflected in the non-incentivized belief elicitation about business success (Table D.5). Female businesses are predicted to have failed with a 4.08 percentage points (Column 1, $p = 0.063$) higher probability than an otherwise identical male business, corresponding to 18.2 percent of the average for a male business (22.5 percent). Seventy percent of this effect stems from loan officers predicting business failure rather than small profits, although this reduction is not statistically significant on its own.

After the initial investment decision, loan officers have the possibility to acquire additional information about the payoff-relevant pitch deck and to reconsider their investment choice. While the point estimate for requesting additional information is positive for female businesses (Table 4.2), it is statistically insignificant. The lack of a significant effect prevents us from concluding whether the effect is driven by statistical discrimination (Phelps, 1972; Arrow, 1973). Nevertheless, the null effect of founders' gender in loan officers' assessment of the quality of the research idea in Table D.6 rather speaks against taste-based discrimination driving the effect (Becker, 1957). The null result on idea quality also supports the hypothesis that the observed bias is rather driven by loan officers' beliefs about gender differences in implementation ability or implementation constraints female-led businesses face. The idea quality measure is not incentivized and its validity may be limited in contrast to the incentivized measures. Yet it correlates with the incentivized measures in the expected direction (Appendix Table D.7).

Table 4.2: Information Request [Individual Entrepreneurs]

	1[Request additional information]								
	Requested info (1)	# items (2)	Team Member (3)	References (4)	Experience (5)	Network (6)	Family F. (7)	External F. (8)	Sales (9)
Female Entrepreneur	.061	.043	-.009	-.027	-.007	-.029	.040	.062	.011
	(.104)	(.264)	(.064)	(.041)	(.056)	(.052)	(.055)	(.058)	(.073)
Mean Dep. Var.	.37	.72	.12	.1	.1	.097	.046	.097	.15
Observations	86	86	86	86	86	86	86	86	86

Notes. OLS regressions of the decision to request additional information about the pitch deck on the gender of the individual entrepreneur who founded and implemented the business. Column (1) reports the binary option of whether the respondent decided to request additional information or not. Column (2) reports the total number of information items requested by the respondent. Columns (3)-(9) report the results for different information items, including: *all team members owning the business, professional references for business owners, professional experience of the business owners, professional network of the creators, financial support from family members received by this business, external financing obtained, volume of sales, revenues, profit margins, and none of the above. Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We also document substantial heterogeneity in the magnitude of the bias by loan officer characteristics. First, the bias against female businesses using incentivized outcomes is most pronounced among loan officers who also exhibit greater general gender bias as measured using aggregated and averaged responses to three selected International Social Survey Programme questions on gender norms (Table 4.1, Panels A and B, Columns 2-3) (Scholz et al., 2014). While we do not find support for a general animus against female-led businesses, this finding indicates that the observed gender bias is reinforced by general gender biased stereotypes: Loan officers exhibiting higher general gender bias may hold stronger beliefs about gender differences in business implementation.[16]

Second, the effect is concentrated among female loan officers, even though for this sub-sample the effect is only statistically significant for the best business choice (Table Table 4.1, Panels A and B, Columns 4-5). Even though it is not statistically distinguishable from the insignificant bias of male loan officers, the larger bias of females is in line with the so-called queen bee syndrome (Staines, Tavris, and Jayaratne, 1974).[17] In this syndrome, women who achieve individual success in male-dominated environments and attain high-status positions are more prone to supporting gender stereotypes. While this effect has mostly been used to describe behavior in hierarchical labor market settings, a financial institution may exhibit similar power asymmetries between loan officers and loan applicants.

[16] Gender bias is measured based on three questions, see table notes of Table 4.3. The third question could also be misinterpreted and hence not reflect gender bias properly. Defining gender bias only based on the first two questions, does not change the direction of coefficients and the magnitude of results (Appendix Table D.8).

[17] Similarly, Bagues and Esteve-Volart, 2010 document that female candidates for positions in Spanish civil service are less likely to be hired if a hiring committee has a higher fraction of females.

Third, loan officer experience does not seem to affect the general gender bias but it reduces discrimination at the top, i.e. when selecting the best performing business (Table 4.1, Panels A and B, Columns 6-7). However, given the relatively large variance spanning between very recently employed loan officers and one officer serving 23 years, we see that the bias in selecting best business is only reduced for the above median group with at least six years of experience. The finding should by no means be interpreted causally as we cannot separate experience effects from, for example, selection. However, even if the causal link was the dominant factor, the dampening effect only occurs in the medium to long term and does not lead to the complete elimination of the bias.

The clear pattern documenting gender bias is further supported by the fact that the data is not driven by limited comprehension, by limited effort on the side of the respondents, or by other confounds. The patterns in responses are consistent across different variables capturing project quality. Examining correlations between the investment measure and other outcomes such as our second incentivized measure of selecting the business as best performing, a non-incentivized rating of idea quality, or a probabilistic belief of whether the project was likely to achieve high profits, shows that all the variables are positively correlated, and the Pearson's correlation coefficients are highly statistically significant at $p < 0.01$ (Table D.7). Reassuringly, the probabilistic belief of whether the business has failed is negatively and significantly ($p < 0.01$) correlated with the above discussed variables. Finally, names assigned to the pitch decks were common enough not to be attributable to a specific demographic characteristic, while being clearly gender-specific. This implies that other characteristics are unlikely to confound the discussed gender effect.

In sum, we observe a robust pattern of gender bias disfavoring individual female businesses by Ugandan loan officers. The effect is strongest in individuals exhibiting gender bias in other domains and seems to be stronger for female loan officers. The bias is not related to beliefs about quality of business ideas but is rather driven by differences in beliefs about implementation ability or implementation constraints.

4.3.2 No Gender Bias in the Evaluation of Entrepreneurial Teams

Focusing on teams in which founder and implementer are different individuals, we no longer observe a gender bias. We document no systemic gender bias of loan officers when evaluating teams in incentivized investment decisions. We also establish that the null result is well identified and sufficiently statistically powered, and it is not driven by lack of variance in the data, by limited attention of respondents, or by limited quality of responses.

First, we find no gender difference in the incentivized investment decision of loan officers. Our results imply that loan officers do not expect any gender differences in the ability of teams of entrepreneurs to generate profit. Panel A of Table 4.3 presents the results in a regression analysis. The dependent variable is the amount (in UGX) invested in each pitch deck. The regressions report coefficients for indicators for a pitch deck having a female founder, a female implementer, and a joint female founder and female implementer.[18] The excluded category is a pitch deck with a male founder and a male implementer. Each regression includes individual and pitch deck fixed effects and standard errors are clustered at the individual level. We also report an F-statistic and a p-value of a test of the sum of all three coefficients, in other words comparing the difference between a business with a male founder and implementer to one with a female founder and implementer. Column 1 of Panel A shows that, on average, there is no statistically significant effect of either of the gender combinations. The point estimates are very small, not exceeding two percent of the mean of the dependent variable. We also do not observe any difference in the cumulative distributions of the different founder and implementer combinations (Figure D.2).

[18]The specification presented deviates from the pre-specified specification in the pre-analysis plan. There, we assumed no differences between mixed-gender and same-gender teams. Since this assumption does not hold in the data, we deviated from the pre-specified specification in our analysis. We present the pre-specified specification for the investment decision in Appendix Table D.9 and, with the strong assumptions discussed above, our conclusions for teams remain unchanged. We detail the deviations form the pre-analysis plan in Appendix D.5.

Table 4.3: Investment and Best Business Decision [Teams of Entrepreneurs]

Panel A: Investment		Gender bias		LO gender		Experience	
		Low	High	Female	Male	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Founder	7.27	-179.28	201.81*	-86.33	77.00	-59.94	64.89
	(86.11)	(120.96)	(121.12)	(123.87)	(118.89)	(123.70)	(119.37)
Female Implementer	-59.54	-97.11	-27.13	-188.07	38.94	-110.85	-12.41
	(83.68)	(126.27)	(108.70)	(130.93)	(106.57)	(123.38)	(116.28)
Female Founder&Implementer	40.33	137.75	-54.46	209.18	-93.08	85.75	5.38
	(120.58)	(181.80)	(157.82)	(190.10)	(154.91)	(182.20)	(163.65)
Mean Dep. Var.	3,352.2	3,395.6	3,308.3	3,353.5	3,351.2	3,303.0	3,393.4
F-Statistic	.017	1	.91	.25	.031	.43	.2
P-value	.9	.31	.34	.62	.86	.51	.66
Observations	1804	936	868	804	1000	796	1008
Panel B: Best Business							
Female Founder	-0.01	-0.07*	0.05	-0.06	0.03	-0.00	-0.02
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
Female Implementer	-0.02	-0.04	-0.00	-0.09*	0.03	-0.06	0.01
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Female Founder&Implementer	0.08**	0.14**	0.03	0.18***	0.00	0.08	0.09
	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Mean Dep. Var.	0.21	0.23	0.18	0.27	0.16	0.21	0.21
F-Statistic	2.4	.27	2.5	.43	2.2	.15	3
P-value	.12	.6	.12	.51	.14	.7	.086
Observations	1804	936	868	804	1000	796	1008

Notes. OLS Regressions of the dependent variable *Investment* (Panel A) or *Best Business* (Panel B) on the gender of the founder and the implementer in teams in which these are different individuals. Panel A reports the incentivized decision of how much to invest in the pitch deck business from 0-5,000 UGX. Panel B reports the incentivized decision of which of all the businesses is the best idea of all the ones seen; it is a probability. Column (1) reports the average effect, and Columns (2)-(7) split the observations according to different relevant observable characteristics. (2)-(3) split by gender bias following International Social Survey Programme gender bias metrics. It incorporates three questions: *A man's job is to earn money; a woman's job is to look after the home and family, a job is alright, but what most women really want is home and children,* and *family life suffers when the woman has a full-time job*. The sample is split at the median. (4)-(5) are split according to the self-reported gender of the respondent, and (6)-(7) are split according to the median experience level. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are clustered at the individual level and reported in parentheses. The table includes pitch deck and individual FE in Panel A and pitch deck FE in Panel B. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Examining heterogeneity, we present the result of a regression specification by gender of the loan officer, by gender bias of a loan officer, and by loan officer experience. The null result documented in the aggregate sample holds true also for all sub-group analyses. If anything, we observe a marginally statistically significant positive effect for a female founder and male implementer business relative to a male founder and implementer business among loan officers with high general gender bias ($p = 0.097$), but this does not differ from what would be expected statistically and it is also not robust to a modified measure of the gender bias.[19]

Reassuringly, as in the case of individual entrepreneurs, we document that the result is not driven by limited comprehension or by limited effort on the side of the respondents. First, Table D.10 shows that the cross-correlations across different variables capturing project quality are going in the same direction and are similar to those observed for single businesses (as in Table D.7). All Pearson's correlation coefficients are highly statistically significant at $p < 0.01$.

Second, the investment measure is incentivized, which motivates loan officers to pay attention and to carefully consider their choices. Removing the choices of ten percent of individuals who required most clicks to answer the comprehension questions correctly does not change the results (Appendix Table D.12). Third, the null result is not a product of a lack of variance in the dependent variable. Histograms in Figure D.3 show that there is a sufficient variation, and the distributions are continuous. Fourth, even though all projects we selected are evaluated positively with investments averaging 3,317 UGX, there is a statistically significant variance between investment in the projects with a range of average investments of 3,065 UGX for pitch deck 2 to 3,460 UGX for pitch deck 4 ($p < 0.01$, see Appendix Table D.11). Ceiling or floor effects thus cannot explain the null result, either. Appendix Table D.11 shows that the null result holds across all pitch decks, respectively. It is particularly notable that it holds even for pitch deck 1. This decision is the first investment decision the loan officers made. It is thus closest to a between subject design, which is least susceptible to possible experimenter demand or order effects. Even though we do not randomize the order of pitch decks, pitch deck 1 ranks in the middle of the quality ranking, so it is an unlikely outlier.

In contrast to the gender bias documented for individual entrepreneurs in the previous sub-section, we find no differential effect by gender or gender composition of a team of entrepreneurs on incentivized investment decisions. Our design allows us to conclude that the lack of bias is true both for business founders and implementers. The effect is not caused by lack of variation, low power, lack of attention, or by other confounds. As the investment behavior is correlated with other proxies of business idea quality, we are confident that the measure is valid, and there is indeed no gender bias in these investment decisions.

[19] As in Footnote 16, we define gender bias based on the first two questions. This does not change the direction of coefficients and the magnitude of results, but renders the female founder coefficient insignificant in the case of loan officers with high gender bias (Appendix Table D.8).

4.3.3 Gender Heterogeneity in Entrepreneurial Teams affects Selection of Top-performing Business

In Table 4.3, Panel B, we report effects of pitch deck gender composition on loan officer's propensity to select a pitch deck as the best performing business from the evaluated five pitches. We use the same regression specification as in Panel A, but the dependent variable is now an indicator for a given project being selected as the best performing business. Thus, we no longer include individual fixed effects. Column 1 in Panel B presents aggregate results. It reveals a positive marginal effect of 8 percentage points for the female founder and implementer business ($p = 0.041$). This effect is statistically significant when compared to either of the mixed-gender teams as both are individually statistically insignificant but with a negative point estimate. However, the overall effect of female entrepreneurial teams compared to male ones is marginally insignificant ($p = 0.124$), indicating no strict preference of female teams over male teams.

Interestingly, the pattern of team gender heterogeneity aversion emerges especially when examining subgroups. The heterogeneity aversion is manifesting itself through a similarly sized negative coefficient for female founder and female implementer indicators, together with a positive female founder & implementer indicator, and an insignificant F-test comparing same-gender teams. In other words, while there is no difference in loan officer decisions when evaluating all-female and all-male teams, there is a relative penalty for mixed-gender teams. We document such pattern for loan officers with low gender bias (Column 2, Panel B) and for female loan officers (Column 4, Panel B). Even though the bias against mixed-gender teams is not universal, it is present among a sizable group of loan officers.

Our results are robust to focusing on the more attentive loan officers (Appendix Table D.12). The effect also does not seem to be driven by a specific pitch deck when analyzing the results for each pitch deck separately (Appendix Table D.11).

Overall, while we observe a bias against selecting pitch decks with mixed-gender founder and implementer combinations, the results do not reflect a general bias against a specific gender as in the case of investment decisions for business teams.

We detect the effect for the right tail of the performance distribution of the business and not on the business survival, as proxied by the incentivized investment decision. Reassuringly, even though the effects are insignificant, the point estimates for a non-incentivized belief about the business achieving large profits go in the same direction, while no such effect emerges for beliefs about business failure or the business achieving small profits (Table D.13).

A possible explanation may be that loan officers are relatively less familiar with entrepreneurial teams, and mixed-gender teams even more. Loan officers may then feel less qualified to evaluate such teams or their decisions may be more noisy. Standard models of information processing would predict increased demand for any possible information about businesses that are less familiar, as the informational value of extra information would be higher as long as it is expected to change prior beliefs. Yet we find no effect on requesting additional information (Table 4.4). It is noteworthy that the average demand for information and the number of information requested does not differ across same-gender and mixed-gender teams. This speaks against a possible explanation that the bias is caused by loan officers being less familiar with mixed-gender business teams.

Table 4.4: Information Request [Teams of Entrepreneurs]

	1[Request additional information]								
	Requested info (1)	# items (2)	Team Member (3)	References (4)	Experience (5)	Network (6)	Family F. (7)	External F. (8)	Sales (9)
Female Founder	-.10	.08	.05	-.03	.06	.05	.02	.00	-.07
	(.08)	(.23)	(.05)	(.05)	(.05)	(.04)	(.03)	(.04)	(.06)
Female Implementer	-.09	-.04	-.01	-.04	-.01	.02	.01	.02	-.03
	(.07)	(.19)	(.05)	(.05)	(.04)	(.04)	(.03)	(.04)	(.06)
Female Founder&Implementer	.17	.11	-.02	.01	-.00	-.04	.01	-.01	.15
	(.11)	(.34)	(.07)	(.06)	(.07)	(.06)	(.05)	(.07)	(.09)
Mean Dep. Var.	.39	.75	.13	.094	.097	.083	.055	.1	.19
F-Statistic	.0335	.342	.168	1.03	.82	.382	.934	.126	.364
P-value	.855	.559	.683	.311	.366	.537	.334	.723	.546
Observations	365	365	365	365	365	365	365	365	365

Notes. OLS regressions of the decision to request additional information about the pitch deck on the gender of the founder and the implementer in teams in which these are different individuals. Column (1) reports the binary option of whether the respondent decided to request additional information or not. Column (2) reports the total number of information items requested by the respondent. Columns (3)-(9) report the results for different information items, including: *all team members owning the business, professional references for business owners, professional experience of the business owners, professional network of the creators, financial support from family members received by this business, external financing obtained, volume of sales, revenues, profit margins, and none of the above. Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are reported in parentheses and are clustered at the individual level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

4.3.4 No Difference in the Evaluations of Individuals and Teams

Finally, we show that there is no general difference in loan officers' evaluations of individual entrepreneurs and teams. We have established two key results. First, individual female entrepreneurs are evaluated worse relative to males. Second, we observe no systematic gender bias in entrepreneurial teams. A natural question is how evaluations for teams differ from evaluations of individual entrepreneurs. Table 4.5 answers the question by reporting results of a regression with an indicator for an individual entrepreneur. The omitted variable is an entrepreneurial team. The regressions controls for pitch deck fixed effects and standard errors are clustered at the level of the loan officer. We examine the entire range of outcomes. Columns 1, 2, and 5-7 also control for individual fixed effects.

Table 4.5: Teams vs. Individuals

	Investment (1)	Best Business (2)	Requested Info (3)	# Info (4)	P[failure] (5)	P[small profits] (6)	P[large profits] (7)
Individual	46.62	.04	.04	.07	-.69	.39	.29
	(62.18)	(.03)	(.06)	(.17)	(1.10)	(1.20)	(1.36)
Mean Dep. Var.	3317	.19	.39	.72	25	40	35
Observations	2255	2255	451	451	2255	2255	2255

Notes. OLS Regressions. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Clustered standard errors at the individual level, as well as round fixed effects and individual fixed effects except for Columns (3) & (4). These columns have robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The results show no systematically different treatment of individual entrepreneurs and teams for the investment decision, the probability of picking the business as the best performing one, and all other outcomes of interest. All null results are relatively precisely estimated as being not statistically significantly different from zero. Hence, we do not find any support for entrepreneurial teams being treated differently from individuals.

We conjecture that neither unfamiliarity with entrepreneurial teams, nor a specific preference or a belief about running a business in teams can explain the difference in gender bias documented for individual entrepreneurs and teams. Unfamiliarity would likely result in an increased demand for information and with an increased variance over the profit distribution. Columns 4-8 do not support this hypothesis. Similarly, if loan officers had a specific preference for either individuals or teams, or if their beliefs about the quality of either type of business differed in general, we should observe this by systematically different outcomes. However, we do not detect differences along any of the dimensions of interest.

4.4 Conclusion

In this study, we analyze whether the evaluation of start-up business potential differs by the gender of the entrepreneur, by the team size, and their interaction. To answer this question, we implement a lab-in-the-field experiment in which 451 loan officers in Uganda evaluate real-life business pitch decks. Our experimental design manipulates both dimensions: gender of the entrepreneur(s) and the formation of teams. A novel design separating the business founder from the implementer in our two-person entrepreneurial teams allows us to locate potential bias either in the evaluation of the business idea or the perceived entrepreneurial capabilities and implementation challenges an entrepreneur faces when operating a business.

We find a sizable gender bias in business evaluations of individual entrepreneurs and no such bias for entrepreneurial teams. The bias for individual entrepreneurs is more pronounced among loan officers who hold gender biased attitudes, who are less experienced, and who are female. We do not find gender differences in costly screening effort or in the subjectively assessed quality of the business idea, supporting neither animus against female-developed business ideas, nor perceived lack of information about individual female led enterprises. Hence, we attribute the documented gender bias to differential assessments of individual women's entrepreneurial ability to operate a business and of the external constraints such businesses are facing. On average, teams are not evaluated nor screened differently than individual entrepreneurs. While loan officers show some preference for same-gender teams – regardless of the gender of the entrepreneurs – we do not find any systemic bias against any female entrepreneur on the team. This further argues against animus as a potential driver of gender bias for individual entrepreneurs.

It remains puzzling why only individual female entrepreneurs are facing a penalty, and not female entrepreneurial teams. While our research design does not allow us to provide a definite answer, we speculate that female teams signal greater commitment to the business. Teams of entrepreneurs are less likely to establish businesses out of necessity, a reason for opening a business stated predominantly by women (Kelley, Singer, & Herrington, 2016). It seems that loan officers do not expect such a difference for businesses founded and operated by individual males. In addition, previous literature predominantly reports discrimination against females in male-dominated sectors (Hebert, 2020; Brock & De Haas, 2023). The selected pitch decks in our study are on average rather attributed to male entrepreneurs in our design validation survey: The selected pitch decks turned out to be perceived about 0.2 standard deviations more male ($p = 0.115$) than the not selected ones. Therefore, loan officers may perceive a gender incongruence for female entrepreneurs in the presented businesses. Still, teaming up seems to counteract the higher barriers that women arguably have in these sectors and potentially even signal positive selection of these entrepreneurs (Goldin, 2020; N. Ashraf et al., 2023). Teaming up may also demonstrate cooperativeness, a trait associated with successful entrepreneurship (Cooper & Saral, 2020), and indicate high social capital, which appears to be particularly important to female entrepreneurs (Cohoon, Wadhwa, & Mitchell, 2010).

Regarding the generalizability of our results, we follow the SANS (selection, attrition, naturalness, and scaling) classification of List (2020). On selection, our study has been conducted among almost half of the entire population of loan officers of a major Ugandan bank for entrepreneurial finance, oversampling branches closer to major urban areas. Attrition was not an issue. Only eight loan officers present at the time of the experimental session did not participate due to other commitments. On naturalness, despite the decisions being framed in the context of a research study, loan officers made incentivized decisions in their regular workplace about real start-up pitch decks. The pitch decks were selected from a pool of pitches developed for a competition attended by Ugandan venture capitalists where stakes were high. For the purposes of the study, we only manipulated the names on the pitch decks to signal gender. Finally, we also comment on potential for scaling of our results beyond the sample studied. While Uganda is characterized by rather low financial development (rank 164 out of 183 for the Financial Development Indicator (IMF, 2021)) and gender equality (rank 131 out of 170 of the UN Gender Inequality Index (UNDP, 2022)), our results on gender bias, in particular the facts that they are driven by loan officers with larger gender bias, confirm the results by Brock and De Haas (2023) for Turkey, a country that scores much higher on financial development (rank 38 out of 183) and gender equality (rank 65 out of 170). This suggests that our results are indeed relevant for countries even beyond those similar to Uganda in terms of financial development and gender equality.

Our results have several implications. First, the observed bias against individual female entrepreneurs can be attributed to loan officers' beliefs that women have lower capabilities to run the business and face more pronounced implementation barriers. Understanding whether such beliefs are correct or whether and how they are biased would be critical for designing interventions aimed at reducing the bias (Bohren et al., 2023). In the first case, more tailored policies to reduce structural disadvantages facing individual female entrepreneurs would be required. In the second case, loan officers should be provided with more accurate information regarding individual female entrepreneurs' performance to correct their beliefs. This seems particularly important since biases based on incorrect beliefs reinforce existing gender gaps, impeding a possible corrections of wrong beliefs without an external stimulus.

Second, the results documenting the lack of a bias for entrepreneurial teams, and the equality in average evaluations of teams and individual entrepreneurs introduce more nuance to the discussion about the role of gender in access to finance and firm growth. Since start-ups with teams of entrepreneurs are more profitable relative to individual entrepreneurs in high-income countries (Åstebro & Serrano, 2015) and start-up accelerators and incubators promote team creation, access to finance for team enterprises may not be disadvantageous to women. That is, as long as entrepreneurs can credibly signal the team composition of their business or apply for funding jointly as a team. Moreover, policies aimed at team creation for start-up enterprises may have an additional benefit of equalizing access to finance. In a dynamic setting, the penalty in evaluations of individual female enterprises may also contribute to under-representation of female businesses among larger firms, due to the difficulties at the start of their potential growth trajectory. What remains to be understood is why women face a penalty when running a business individually and not in a team of female entrepreneurs and why Ugandan loan officers do not evaluate teams more favorably than individual entrepreneurs.

Third, the team composition seems to be relevant when assessing the potential business success. Loan officers show a mild bias against mixed-gender teams. Loan officers reported in conversations that they foresee coordination and communication problems in mixed-gender teams. A promising avenue for future research is to study the role of entrepreneurial team composition and its interconnections with access to finance.

A

Appendix to Chapter 1

Time Pressure and Regret in Sequential Search

A.1 Theoretical Search Model

Standard Information Environment

To derive testable behavioral hypotheses for the experimental design, we incorporate regret aversion into one of the most classic and simple search models, building on the formulation of Schunk (2009).[1] In the model, agents have an inelastic demand for one unit of a good, receive offers sequentially, and they incur a (fixed) search cost for every offer that they request. We allow for perfect recall, such that agents can always take the lowest price encountered so far. There is no limit on the number of offers that can be requested and the prices are randomly drawn from a previously known

[1]This relates to other theoretical models that incorporate regret in static frameworks like currency hedging (Michenaud & Solnik, 2008), insurance choices (Braun & Muermann, 2004) or the expansion of the choice set (Irons & Hepburn, 2007). In sequential decisions, general approaches to model dynamic choices under regret (e.g., Krähmer & Stone, 2008) have been applied to investment decisions (Muermann & Volkman, 2007) and asset-selling problems (Strack & Viefers, 2021).

discrete uniform distribution. The distribution function from which the offers are drawn is $F(.)$ with range $[l, h]$. The search costs for each requested offer are denoted as c. Both the distribution function and the search costs are known to the agent. The agent maximizes profits (π), which are calculated as the difference between induced valuation (v) for the good and the costs for the purchase. This cost consists of the total search cost plus the final price to be paid (p). The best price observed so far is denoted by (m_t). Intuitively, to request another offer, the sure loss of c must be outweighed by the possibility of finding a better price in $t+1$.[2]

Payoff-maximizing agent. The optimal behavior for a risk-neutral agent is a constant reservation price strategy (Lippman & McCall, 1976). To calculate this reservation price, it is sufficient for the agent to compare the benefits from stopping the search now and the benefits requesting one additional offer and stopping afterward. This is displayed in Equation A.1.

$$\begin{aligned} \pi(v-m_t) &= [1-F(m_t)]\pi(v-m_t-c) + \int_l^{m_t} \pi(v-x-c)\,\mathrm{d}F(x) \qquad \text{(A.1)} \\ \Leftrightarrow \pi(v-m_t) &= [1-F(m_t)]\pi(v-m_t-c) + \int_l^{m_t-c} \pi(v-x-c)\,\mathrm{d}F(x) \\ &\quad + \int_{m_t-c}^{m_t} \pi(v-x-c)\,\mathrm{d}F(x) \end{aligned}$$

The left-hand side represents the value from stopping the search. The right-hand side is the value from requesting another offer. The first term on the right side corresponds to the cases where no better price is found. The second term in the first line corresponds to prices that are below the current best price (m_t) and weights the resulting profits by their probability. Given the parametrization in the experiment (v=50; discrete uniform distribution with range $[1, 100]$), we solve

$$\begin{aligned} 50-m_t &= \frac{100+c-m_t}{100}\cdot(50-c-m_t) + \sum_{i=1}^{10}(50-c-x)\cdot\frac{1}{100} \\ &= \frac{100+c-m_t}{100}\cdot(50-c-m_t) + \frac{m_t-c}{100}\cdot\frac{99-m_t-c}{2} \end{aligned}$$

for the reservation price m_t which equals the benefits from stopping the search now with the benefits of requesting one additional offer. For example, in the case of search

[2]We refer to every decision between stopping at offer t or requesting offer $t+1$ as a round, meaning that every search task k consists of up to 25 rounds.

cost $c = 2$, this results in a reservation price of 20.56.

In the second row of Equation A.1, we distinguish between the cases where better prices outweigh the search costs ($m_t - m_{t+1} > c$), and the cases where they do not. This allows us to draw a comparison with the optimization problem of a regret-sensitive agent.

Regret-sensitive agent. We also derive predictions for a regret-sensitive agent (Bell, 1982; Loomes & Sugden, 1982). We make the simplifying assumption that regret is a function of the difference between the payoffs of the chosen and the unchosen option. Accordingly, the utility from choosing option i over k under the state of the world j is defined as: $m_{ij}^k = \pi(x_{ij}) - R[\pi(x_{kj}) - \pi(x_{ij})]$. The agent both derives utility from the material benefits from the choice of i, but also from the comparison of the chosen and the unchosen option. The regret/rejoice-function R specifies how much the comparison of actual and counterfactual outcomes affects the individual's utility. As common (e.g Zeelenberg, 1999; Muermann & Volkman, 2007; Michenaud & Solnik, 2008), we build on the observation that regret is felt more intensely than rejoice (Bleichrodt, Cillo, & Diecidue, 2010). For simplicity, we assume that the agent does not experience (and anticipate) any rejoice. The agent experiences negative utility if the unchosen option had led to higher payoffs. Conversely, the agent does not experience positive utility if the chosen alternative led to higher profits. We assume regret aversion; that is, an increasing convex R in the positive domain of regret.

The experience and anticipation of *inaction regret* induce the two commonly observed anomalies in standard search tasks: early stopping and the recall of previously rejected offers. The utility from stopping at a lowest price m_t in round t becomes $u(m_t) = \pi(v - m_t) - R(\pi_{max_t} - \pi_t)$. Regret is defined as a function of the foregone profits by not having stopped at the payoff-maximizing offers up to t. π_{max_t} denotes the payoffs at the ex-post optimal stopping point. This maximum serves as a reference point for the feelings of regret. π_t denotes the payoff from stopping in round t.

We incorporate *inaction regret* into Equation A.1. Equation A.2 models optimal decision making for regret-sensitive agents using one-step forward-induction. Current feelings of regret enter on the left-hand side, anticipated feelings on the right-hand side. On the right-hand side, the first term captures the case where the next draw does not yield a better price than m_t. The second term describes the situations in which a payoff-increasing price was drawn. The third term corresponds to prices that are better than m_t, but do not outweigh the search costs (c).

$$\begin{aligned}\pi(v-m_t)-R(\pi_{max_t}-\pi_{m_t}) =& [1-F(m_t)][\pi(v-m_t-c)-R(\pi_{max_t}-\pi_t-c)] \\ &+\int_{l}^{m_t-c} \pi(v-x-c)\,\mathrm{d}F(x) \\ &+\int_{m_t-c}^{m_t} [\pi(v-x-c)-R(\pi_x-\pi_{max_t}-c)]\,\mathrm{d}F(x)\end{aligned} \tag{A.2}$$

Why would a regret-averse agent search shorter than an expected profit-maximizing individual? In the standard information environment, no feedback about foregone options after stopping is revealed. You only feel regret if you have searched for too long (*inaction regret*). At each decision node, the experience of (additional) regret can occur only by continuing, not by stopping. Accordingly, regret-averse agents have a higher reservation price and therefore request fewer offers. For simplicity, we assume that the current price is the best offer so far. Given $\pi_{max_t} = \pi_t$, the left hand sides of Equations A.1 and A.2 are the same. Nevertheless, the expected value from continuing the search is strictly lower for regret-averse agents. If no better price is found, then not only does the material loss of c reduce utility but so does the regret of not stopping in the previous round. As the continuation value is lower, a regret-averse agent stops searching at a higher price than a pure payoff-maximizer due to the anticipation of (potential) *inaction regret*.

We illustrate the higher reservation price of regret-sensitive agents with the parameters of our experimental design. We assume that the decision-maker receives an initial offer of $m_1 = 22$ and faces search cost of $c = 2$. As illustrated above, a payoff-maximizing agent would continue the search as the offer is above the reservation price. The regret-sensitive decision maker also anticipates aversive feelings of size $\frac{R(1)}{100} + \frac{R(2)*78}{100}$ if they continue the search without encountering a more favorable offer. The first term corresponds to the case in which the next offer m_2 is equal to 21, the second term to cases where the next offer is weakly higher than the current one ($m_2 \geq 22$). Whether to stop the search at $m_1 = 22$ depends on the relative importance of anticipated regret. Assuming the regret function takes the following functional form $R(\pi_{max_t}-\pi_t) = \rho(\pi_{max_t}-\pi_t)^2$, an agent would only continue the search for $\rho < 0.604$. If the sensitivity to feelings of regret is larger, the decision maker stops the search at $m_1 = 22$.

Why would regret-averse agents sometimes exercise recall? A regret-averse agent may use the recall option to avoid additional *inaction regret*. Suppose that a regret-averse agent rationally chose to continue searching in round t and does not find a better price in the subsequent round. Now they experience regret $R(c)$ and anticipate that not finding a better price in the next round leads to $R(2c)$. Because the regret function is convex, the (potential) increase in aversive feelings of regret is higher in this decision than in the previous decision. This may translate into a higher reservation price and a reversal of the choice to continue the search.

Post-purchase Information Environment

While seeing subsequent prices does not alter the utility function of pure payoff-maximizers, regret-sensitive agents are affected by this variation. Seeing subsequent prices may lead to *action regret*. Participants may blame themselves for having stopped too early when continuing the search would have yielded a higher payoff.[3] Thus, seeing subsequent prices directly affects the utility from stopping and enters the left-hand side of Equation A.2. For simplification, we assume that the agent encountered the best draw in round t. We also ignore *inaction regret* because it is constant across conditions and enters the utility function independently.

The (expected) utility from stopping the search in round t while anticipating to see the next draw in case of stopping becomes $\pi(v - m_t) - \int_l^{m_t - c} R(m_t - c - x)\,\mathrm{d}F(x)$. The second term captures that regret is experienced when the price of the next draw (x_{t+1}) is lower than the previously best price m_t and also compensates for the search cost. If one anticipates seeing all of the draws, then the feelings of regret add up to $\sum_{n=1}^{\infty} \int_l^{m_t - nc} R(m_t - nc - x)\mathrm{d}F(x)$, n denoting the (future) draws.[4]

For a regret-averse agent, the expected utility from stopping the search in t is strictly lower when additional draws are revealed after the end of the search. An agent who solves the problem based on one-step forward-induction anticipates that the same holds when stopping the search after requesting another offer $(t + 1)$. To avoid additional subscripts, the next offer x_{t+1} is denoted as z in the following optimization

[3]This entails the implicit assumption that the agent needs to see the price realization to experience *action regret* (or not), instead of incorporating expectation-based regret (without ever knowing the realization) into every decision.

[4]The upper limit of the integral changes because the likelihood of finding a more favorable offer decreases in each round as it has to compensate for all additional search costs. This is not necessary when defining R only in the positive domain. To allow for a more general definition of R, we maintain this notation. An alternative approach would be to define regret only with respect to the best forgone option. While possible, calculating the probabilities of each regret level conditional on being the highest would have been more complicated.

problem with *action regret*.

$$\pi(v-m_t)-\sum_{n=1}^{\infty}\int_{l}^{m_t-nc} R(m_t-nc-x)\mathrm{d}F(x)=$$
$$[1-F(m_t)][\pi(v-m_t-c)-\sum_{n=2}^{\infty}\int_{l}^{m_t-nc} R(m_t-nc-x)] \tag{A.3}$$
$$+\int_{l}^{m_t}[\pi(v-z-c)-\sum_{n=2}^{\infty}\int_{l}^{z-(n-1)c} R(z-(n-1)c-x)]\,\mathrm{d}F(x)$$

If the next draw does not yield a better price, then the probability of experiencing *action regret* when stopping the search in $t+1$ is lower than in t. This happens because future offers must also compensate for the additional search costs incurred to be advantageous. If a better offer is found in $t+1$, then the expectation of regretting the purchase at the new price is lower because it becomes less likely that future draws will yield a better payoff. Therefore, the variation in the information structure increases the (relative) attractiveness of requesting another offer and induces longer search durations for regret-sensitive agents.

Previous Regret Experience and Urgency: Linking Experimental Design and Theoretical Model

In the experimental design, we go beyond the stylized one-period model outlined so far. We allow for the experience of regret in a previous task as participants face multiple search tasks. We hypothesize that the experience of regret in task k intensifies the anticipation of regret in task $k+1$. As a consequence, experiencing (inaction) regret due to requesting too many offers in task k translates into shorter search in the next task. Experiencing (action) regret due to requesting too few offers in task t translates into longer search in task $t+1$. In our model, this is both consistent with a payoff-maximizing agent becoming regret-sensitive through experience (extensive margin effect) and with the functional form of the regret function R being subject to regret experiences (intensive margin effect).

Our design also takes into account urgency, which our model does not. We can formally link urgency and regret through the introduction of cognitive capacities if we assume that the anticipation of regret depends on the amount of available cognitive resources. A straightforward approach would be to think about cognitive capacities

($\lambda \in [0,1]$) as a scaling factor for regret. The perceived utility from stopping at a lowest price m_t in round t became $u(m_t) = \pi(v - m_t) - \lambda R(\pi_{max_t} - \pi_t)$. For example, if the agent does not have any cognitive resources available ($\lambda = 0$), there is no anticipation of regret. Hence, the reduction of cognitive resources during the decision-making process through time pressure makes the agent less sensitive to feelings of regret. The impact of the regret manipulation on search length is therefore expected to be smaller in treatments with high levels of perceived urgency.[5]

A.2 Additional Figures and Tables

Table A.1: Average Decision Times per Task across Time Pressure Conditions by Feedback Condition

	No-Info			*Info*		
Task	*Low-TP*	*High-TP*	p-value	*Low-TP*	*High-TP*	p-value
1	10.86	5.87	<0.001	11.12	4.46	<0.001
2	9.24	3.50	<0.001	10.82	3.06	<0.001
3	8.54	2.74	<0.001	10.97	2.56	<0.001
4	5.84	2.28	<0.001	7.84	2.49	<0.001
5	5.74	2.20	<0.001	6.01	2.19	<0.001
6	5.63	2.17	<0.001	6.50	2.16	<0.001
7	5.26	2.38	<0.001	5.59	2.33	<0.001
8	6.27	2.38	<0.001	5.08	2.37	<0.001
9	8.60	3.36	<0.001	5.37	2.93	<0.001
10	10.03	3.33	<0.001	6.83	3.14	<0.001

The table shows the average decision times across the time pressure conditions by feedback condition. The p-values are based on non-parametric Mann-Whitney U tests (MWU) on whether the participants' average decision times per task and feedback condition in *Low-TP* and *High-TP* come from the same underlying distribution.

[5]This modeling approach would yield a directed hypothesis on the effect of urgency in environments without post-purchase information. With urgency, we should observe longer (and more efficient) search. At the same time, we acknowledge the multiple channels through which urgency may impact search behavior and do not specify a directed hypothesis in the main text.

Table A.2: Experienced Regret

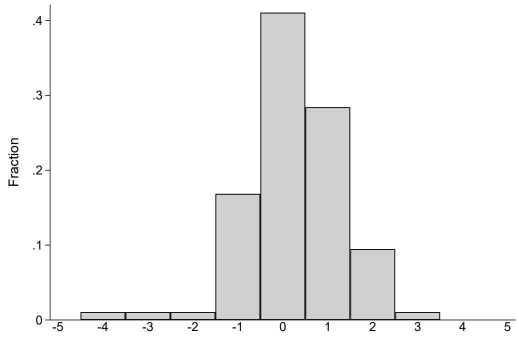
	Number of offers		
	Full Sample	No-Info	Info
	(1)	(2)	(3)
Treatments			
High-TP	.432	.415	.301
	[-.103,.967]	[-.152,.981]	[-.269,.872]
Info	-.115		
	[-.610,.380]		
High-TP X Info	-.135		
	[-.771,.500]		
(Experienced) Inaction Regret	.470	.413	-.695*
	[-.228,1.168]	[-.392,1.218]	[-1.408,.019]
Inaction Regret X Info	-1.082***		
	[-1.815,-.349]		
Inaction Regret X High-TP	-.104	.071	-.077
	[-.858,.650]	[-1.046,1.188]	[-1.042,.888]
(Experienced) Action Regret	-.124	-.121	1.029*
	[-.806,.558]	[-.892,.651]	[-.063,2.121]
Action Regret X Info	1.058**		
	[.225,1.891]		
Action Regret X High-TP	-.757*	-.798	-.772
	[-1.598,.084]	[-1.889,.293]	[-2.062,.518]
# Tasks encountered	.064**	.049	.079**
	[.008,.119]	[-.033,.131]	[.002,.155]
Risk Aversion	-.065	.021	-.123**
	[-.146,.015]	[-.087,.129]	[-.234,-.012]
Loss Aversion	.054	.011	.159
	[-.075,.183]	[-.151,.174]	[-.033,.351]
Constant	4.667***	5.684***	3.469***
	[3.229,6.106]	[3.791,7.576]	[1.836,5.101]
Socio-demographic controls	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes
Observations	1719	855	864

OLS Regressions.*** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. Values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching.Columns (1)-(4) display search behavior in tasks 2-10 and investigate the effect of regret experienced in the previous task. All columns include socio-demographic controls (gender, age, cognitive ability) and price sequence group fixed effects. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 2-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

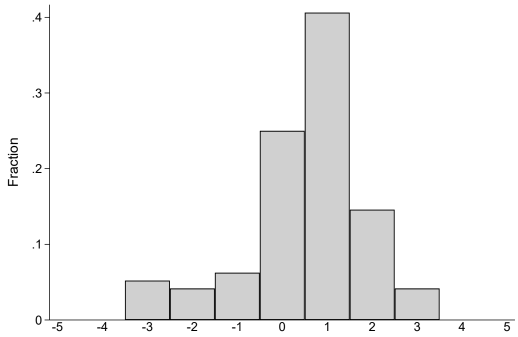
Table A.3: Experienced Regret: Optimality of Search

	Forgone Profits	Optimal	Too few offers	Too many offer
	(1)	(2)	(3)	(4)
Treatments				
High-TP	.971	-.087**	.006	.081**
	[-.271,2.213]	[-.158,-.017]	[-.051,.064]	[.011,.152]
Info	.173	.016	.006	-.022
	[-1.261,1.607]	[-.056,.087]	[-.054,.066]	[-.089,.045]
High-TP X Info	-.606	.008	.009	-.017
	[-2.874,1.662]	[-.094,.109]	[-.072,.091]	[-.110,.075]
(Experienced) Inaction Regret	1.915**	-.124***	-.013	.137***
	[.399,3.430]	[-.215,-.033]	[-.080,.054]	[.057,.217]
Inaction Regret X Info	-1.505	.070	.019	-.089*
	[-3.502,.492]	[-.048,.188]	[-.077,.115]	[-.187,.010]
(Experienced) Action Regret	-1.136	.078**	.002	-.081***
	[-2.669,.397]	[.006,.151]	[-.068,.072]	[-.131,-.030]
Action Regret X Info	3.526**	-.108**	.006	.102***
	[.851,6.200]	[-.215,-.000]	[-.089,.101]	[.026,.178]
# Tasks encountered	-.200**	.023***	-.031***	.008**
	[-.371,-.029]	[.015,.031]	[-.038,-.024]	[.001,.015]
Risk Aversion	-.160	.009	.002	-.011*
	[-.569,.250]	[-.005,.022]	[-.009,.013]	[-.022,.000]
Loss Aversion	-.167	-.010	-.005	.015
	[-.640,.307]	[-.030,.011]	[-.020,.010]	[-.005,.035]
Constant	6.334**	.624***	.413***	-.037
	[.324,12.344]	[.409,.840]	[.214,.612]	[-.214,.140]
Socio-demographic controls	Yes	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes	Yes
Observations	1719	1719	1719	1719

OLS Regressions.*** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. Columns (1) shows an OLS regression, estimating the forgone profits compared to the ex-ante optimal benchmark. Column (2) estimates the likelihood that search behavior was optimal (compared to the ex-ante optimal benchmark) with a (binary) OLS regression. The (binary) dependent variable takes the value 1 if the participant requested the optimal number of offers in the task and 0 otherwise. Column (3) shows the corresponding analysis with the dependent variable taking the value 1 if too few offers were requested and 0 otherwise. In Column (4), the dependent variable takes the value 1 if too many offers were requested and 0 otherwise. All columns refer to search behavior in tasks 2-10. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.



(a) *Low-TP*



(b) *High-TP*

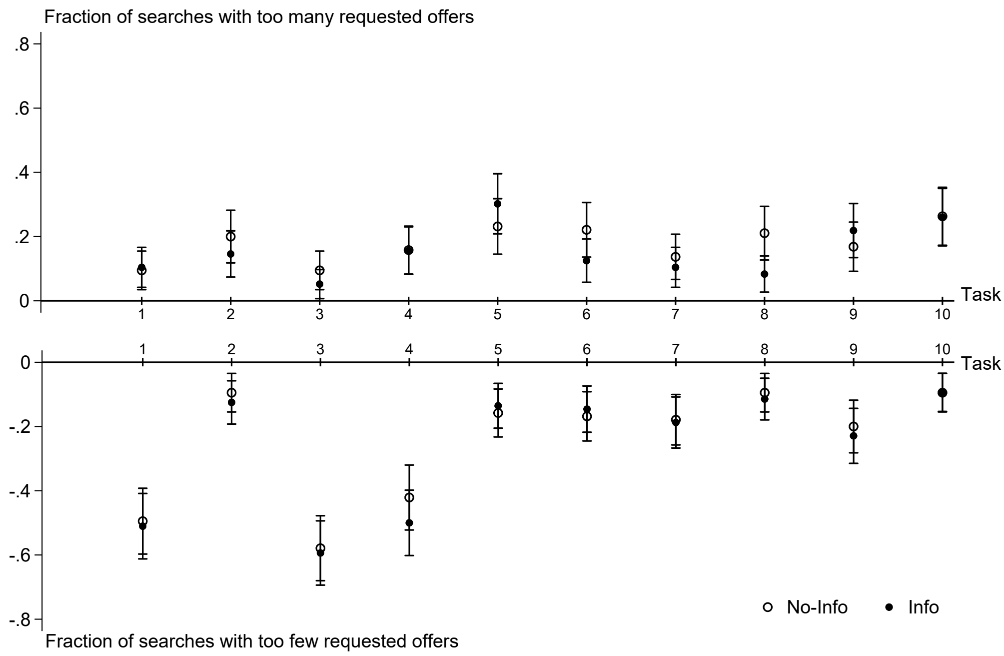
Notes. The figure shows the perceived advantage of having 60 sec for each decision. Positive values indicate that the participant expected to perform better with 60 seconds than with 4 seconds. For example, a value of 1 in the left-hand panel (*Low-TP*) means that a participant expects to have scored one rank lower in the group of six if they had only had 4 seconds. In the right-hand panel (*High-TP*), a value of 1 means that a participant expects to have scored one rank higher in the group of six if they had had 60 seconds.

Figure A.1: Perceived Advantage of 60 Seconds for the Decision by Time Pressure Condition

Table A.4: Probit Regression: Stopping the Search

	1[Stopped Search]			
	(1)	(2)	(3)	(4)
Treatments				
High-TP	-.011	-.007	.148***	.156***
	[-.049,.028]	[-.043,.028]	[.048,.249]	[.057,.256]
Info	-.001	.001	.041	.046
	[-.037,.035]	[-.032,.034]	[-.033,.115]	[-.029,.122]
High-TP X Info	.016	.012	-.132*	-.137*
	[-.039,.071]	[-.039,.063]	[-.277,.013]	[-.274,.000]
# Tasks encountered	-.006***	-.005***		
	[-.009,-.003]	[-.009,-.002]		
Price	-.009***	-.009***	-.010***	-.010***
	[-.010,-.008]	[-.010,-.008]	[-.011,-.008]	[-.012,-.008]
Risk Aversion	.003	.004	.010	.009
	[-.005,.011]	[-.003,.012]	[-.012,.032]	[-.012,.031]
Loss Aversion	-.003	.001	.027*	.032**
	[-.014,.008]	[-.009,.011]	[-.004,.058]	[.001,.063]
Socio-demographic controls	Yes	Yes	Yes	Yes
Price Sequence Group FE	No	Yes	No	Yes
Observations (# of choices)	7226	7226	622	622

Probit Regression.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. The table shows marginal effects at the mean from a probit regression. Columns (1) & (2) display search behavior across tasks 1-10, columns (3) & (4) in task 1. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Price* is the price of the current offer [1,100] the participant faces. *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.



Notes. The upper panel displays the fraction of searches per task in which too many offers were requested. The lower panel shows the fraction of searches, where too few offers were requested. Larger (absolute) values correspond to higher deviations from optimal search behavior.

Figure A.2: Deviation from Optimal Behavior across Tasks by Feedback Condition

A.3 Tasks 2-10

In tasks 2-10, the participants stop on average after seeing 3.84 offers, which are significantly fewer offers compared to the (ex-ante) optimal strategy of an expected payoff-maximizer, requesting 4.47 offers on average ($p < 0.001$, Wilcoxon signed-ranks test). The number of requested offers is very similar across treatments. Search length neither differs between *High-TP* and *Low-TP* ($p = 0.589$; MWU) nor between *No-Info* and *Info* ($p = 0.714$; MWU). This holds equally true when comparing treatments individually and when re-calculating the main regression outcomes for the tasks 2-10 (see Table A.5).

Table A.5: OLS Regression Search Length (Task 2-10)

	Number of offers (Task 2-10)		
	(1)	(2)	(3)
Treatments			
High-TP	.133	.200	.200
	[-.403,.669]	[-.327,.726]	[-.292,.693]
Info	-.059	-.026	-.041
	[-.573,.455]	[-.523,.471]	[-.494,.411]
High-TP X Info	-.138	-.178	-.175
	[-.854,.579]	[-.888,.531]	[-.819,.470]
# Tasks encountered	.064**	.064**	.064**
	[.009,.120]	[.009,.120]	[.009,.120]
Risk Aversion		-.041	-.066*
		[-.122,.040]	[-.144,.012]
Loss Aversion		.048	.045
		[-.095,.190]	[-.084,.174]
Constant	3.453***	4.229***	4.857***
	[2.962,3.945]	[3.139,5.319]	[3.512,6.202]
Socio-demographic controls	No	Yes	Yes
Price Sequence Group FE	No	No	Yes
Observations	1719	1719	1719

OLS Regressions.*** $p < 0.01$, **$p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching. Columns (1)-(3) display search behavior across tasks 2-10. Column (2) adds socio-demographic controls (gender, age, cognitive ability) elicited after all search tasks; Column (3) and (6) additionally include price sequence group fixed effects. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 2-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

A.4 Robustness Checks

A.4.1 Inclusion of Unresponsive Participant

We show that our main regression analyses (Tables 1.3 and 1.4) are robust to the inclusion of one participant who was unresponsive to the price offers from task 3 onward.

Table A.6: OLS Regression Search Length (Unresponsive)

	Number of offers					
	Task 1-10			Task 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments						
High-TP	-.321	-.290	-.275	-.937**	-1.052***	-1.087***
	[-1.146,.505]	[-1.135,.555]	[-1.052,.502]	[-1.695,-.180]	[-1.849,-.255]	[-1.779,-.395]
Info	-.429	-.430	-.433	-.292	-.185	-.203
	[-1.255,.397]	[-1.311,.451]	[-1.249,.382]	[-1.147,.563]	[-1.065,.695]	[-.838,.432]
High-TP X Info	.310	.284	.271	.875	.938	.958
	[-.638,1.258]	[-.666,1.235]	[-.600,1.143]	[-.410,2.160]	[-.361,2.236]	[-.214,2.131]
# Tasks encountered	.090***	.090***	.090***			
	[.039,.141]	[.039,.141]	[.039,.141]			
Risk Aversion		-.059	-.088*		.004	-.079
		[-.151,.033]	[-.177,.001]		[-.174,.182]	[-.255,.097]
Loss Aversion		.034	.008		-.254*	-.232*
		[-.099,.167]	[-.114,.130]		[-.530,.022]	[-.470,.005]
Constant	3.671***	4.616***	6.054***	3.646***	5.721***	4.741***
	[2.995,4.347]	[3.452,5.781]	[3.312,8.796]	[3.054,4.237]	[3.396,8.047]	[2.271,7.211]
Socio-demographic controls	No	Yes	Yes	No	Yes	Yes
Price Sequence Group FE	No	No	Yes	No	No	Yes
Observations	1920	1920	1920	192	192	192

OLS Regressions.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching. Columns (1)-(3) display search behavior across tasks 1-10, columns (4)-(6) for the first task. Columns (1) and (4) show the effect of the treatments. Columns (2) and (5) add socio-demographic controls (gender, age, cognitive ability) elicited after all search tasks; columns (3) and (6) additionally include price sequence group fixed effects. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

Table A.7: Experienced Regret (Unresponsive)

	Number of offers			
	(1)	(2)	(3)	(4)
Treatments				
High-TP	-.190	-.196	-.174	-.185
	[-1.045,.666]	[-1.018,.626]	[-1.010,.662]	[-.988,.618]
Info	-.447	-.068	-.833*	-.436
	[-1.313,.419]	[-.745,.608]	[-1.812,.145]	[-1.152,.280]
High-TP X Info	.180	.238	.171	.214
	[-.732,1.092]	[-.681,1.157]	[-.769,1.111]	[-.699,1.128]
(Experienced) Inaction Regret	.316	1.178*		1.099
	[-.519,1.151]	[-.224,2.581]		[-.264,2.461]
Inaction Regret X Info		-1.790**		-1.715**
		[-3.189,-.390]		[-3.075,-.356]
(Experienced) Action Regret			-.913**	-.816**
			[-1.609,-.218]	[-1.442,-.189]
Action Regret X Info			1.397***	1.305***
			[.477,2.317]	[.426,2.183]
# Tasks encountered	.071**	.075***	.067**	.070**
	[.015,.127]	[.019,.131]	[.011,.122]	[.016,.124]
Risk Aversion	-.087*	-.084*	-.092**	-.087*
	[-.175,.002]	[-.172,.004]	[-.183,-.001]	[-.175,.000]
Loss Aversion	.030	.035	.037	.038
	[-.102,.162]	[-.093,.164]	[-.093,.167]	[-.088,.164]
Constant	6.299***	5.958***	6.572***	6.202***
	[3.272,9.327]	[3.162,8.754]	[3.373,9.771]	[3.319,9.085]
Socio-demographic controls	Yes	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes	Yes
Observations	1728	1728	1728	1728

OLS Regressions.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching. Columns (1)-(4) display search behavior in tasks 2-10 and investigate the effect of regret experienced in the previous task. All columns include socio-demographic controls (gender, age, cognitive ability) and price sequence group fixed effects. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 2-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

A.4.2 Truncated Poisson Regressions

We show that our main regression analyses (Tables 1.3 and 1.4) are robust to a truncated Poisson specification.

Table A.8: Poisson Regression: Search Length

	Number of offers					
	Task 1-10			Task 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments						
High-TP	.006	.021	.021	-.366**	-.400***	-.425***
	[-.131,.144]	[-.114,.156]	[-.106,.149]	[-.652,-.080]	[-.691,-.109]	[-.678,-.172]
Info	-.025	-.013	-.014	-.105	-.062	-.100
	[-.166,.116]	[-.149,.123]	[-.135,.106]	[-.381,.170]	[-.345,.222]	[-.298,.098]
High-TP X Info	-.010	-.019	-.020	.344	.367	.401*
	[-.204,.185]	[-.212,.175]	[-.195,.155]	[-.115,.802]	[-.086,.821]	[-.007,.810]
# Tasks encountered	.023***	.023***	.023***			
	[.010,.036]	[.010,.036]	[.010,.037]			
Risk Aversion		-.011	-.020*		.006	-.025
		[-.033,.012]	[-.043,.004]		[-.056,.068]	[-.088,.039]
Loss Aversion		.005	.005		-.088*	-.089**
		[-.032,.042]	[-.030,.040]		[-.177,.000]	[-.164,-.013]
Constant	1.188***	1.475***	1.580***	1.275***	1.940***	1.587***
	[1.067,1.308]	[1.157,1.794]	[1.232,1.928]	[1.096,1.454]	[1.126,2.754]	[.430,2.743]
Socio-demographic controls	No	Yes	Yes	No	Yes	Yes
Price Sequence Group FE	No	No	Yes	No	No	Yes
Observations	1910	1910	1910	191	191	191

Truncated Poisson Regressions.*** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching. Columns (1)-(3) display search behavior across tasks 1-10, columns (4)-(6) for the first task. Columns (1) and (4) show the effect of the treatments. Columns (2) and (5) add socio-demographic controls (gender, age, cognitive ability) elicited after all search tasks; columns (3) and (6) additionally include price sequence group fixed effects. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

Table A.9: Poisson Regression: Experienced Regret

	Number of offers			
	(1)	(2)	(3)	(4)
Treatments				
High-TP	.058	.054	.060	.057
	[-.080,.196]	[-.077,.185]	[-.071,.192]	[-.070,.184]
Info	-.009	.057	-.093	-.028
	[-.140,.122]	[-.080,.193]	[-.233,.047]	[-.170,.113]
High-TP X Info	-.049	-.035	-.057	-.043
	[-.233,.135]	[-.218,.147]	[-.241,.127]	[-.226,.141]
(Experienced) Inaction Regret	-.024	.126		.114
	[-.143,.095]	[-.029,.282]		[-.037,.265]
Inaction Regret X Info		-.328***		-.317***
		[-.537,-.118]		[-.521,-.113]
(Experienced) Action Regret			-.167**	-.158*
			[-.331,-.003]	[-.320,.005]
Action Regret X Info			.317***	.310**
			[.077,.558]	[.073,.547]
# Tasks encountered	.019**	.020**	.018**	.019**
	[.003,.035]	[.004,.036]	[.002,.033]	[.003,.035]
Risk Aversion	-.019	-.018	-.020*	-.019
	[-.042,.004]	[-.041,.005]	[-.043,.003]	[-.042,.004]
Loss Aversion	.013	.014	.014	.015
	[-.025,.051]	[-.023,.051]	[-.023,.051]	[-.022,.052]
Constant	1.592***	1.535***	1.635***	1.576***
	[1.219,1.965]	[1.158,1.911]	[1.250,2.021]	[1.183,1.968]
Socio-demographic controls	Yes	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes	Yes
Observations	1719	1719	1719	1719

Truncated Poisson Regressions.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching. Columns (1)-(4) display search behavior in tasks 2-10 and investigate the effect of regret experienced in the previous task. All columns include socio-demographic controls (gender, age, cognitive ability) and price sequence group fixed effects. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 2-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

A.4.3 No Switchpoint

We show that our main regression analyses (Tables 1.3 and 1.4) are robust to controlling for risk attitudes and loss attitudes without by calculating the number of safe choices instead of a switchpoint.

Table A.10: Search Length (No Switchpoint)

	Number of offers					
	Task 1-10			Task 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments						
High-TP	.022	.098	.108	-.973**	-1.090***	-1.066***
	[-.461,.506]	[-.378,.575]	[-.340,.555]	[-1.737,-.208]	[-1.906,-.274]	[-1.764,-.368]
Info	-.086	-.041	-.056	-.327	-.188	-.190
	[-.571,.399]	[-.512,.430]	[-.471,.360]	[-1.188,.534]	[-1.079,.702]	[-.838,.457]
High-TP X Info	-.033	-.075	-.073	.910	.988	.978
	[-.704,.639]	[-.741,.590]	[-.672,.526]	[-.379,2.199]	[-.329,2.306]	[-.201,2.158]
# Tasks encountered	.079***	.079***	.079***			
	[.032,.125]	[.032,.125]	[.032,.125]			
Risk Aversion		-.044	-.077		-.018	-.105
		[-.142,.054]	[-.169,.016]		[-.226,.189]	[-.314,.104]
Loss Aversion		.044	.047		-.158	-.104
		[-.104,.192]	[-.078,.171]		[-.462,.146]	[-.358,.150]
Constant	3.391***	4.251***	4.691***	3.681***	5.324***	4.157***
	[2.988,3.793]	[3.280,5.223]	[3.500,5.883]	[3.081,4.281]	[3.190,7.457]	[1.770,6.545]
Socio-demographic controls	No	Yes	Yes	No	Yes	Yes
Price Sequence Group FE	No	No	Yes	No	No	Yes
Observations	1910	1910	1910	191	191	191

OLS Regressions.*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching.Columns (1)-(3) display search behavior across tasks 1-10, columns (4)-(6) for the first task. Columns (1) and (4) show the effect of the treatments. Columns (2) and (5) add socio-demographic controls (gender, age, cognitive ability) elicited after all search tasks; columns (3) and (6) additionally include price sequence group fixed effects. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as the number of safe choices, as described in Footnote 19.

Table A.11: Experienced Regret (No Switchpoint)

	Number of offers			
	(1)	(2)	(3)	(4)
Treatments				
High-TP	.241	.230	.252	.245
	[-.256,.738]	[-.239,.700]	[-.224,.727]	[-.213,.703]
Info	-.043	.192	-.335	-.105
	[-.495,.410]	[-.282,.666]	[-.827,.156]	[-.600,.390]
High-TP X Info	-.188	-.146	-.217	-.173
	[-.833,.457]	[-.781,.490]	[-.860,.426]	[-.814,.469]
(Experienced) Inaction Regret	-.080	.478		.425
	[-.494,.334]	[-.121,1.077]		[-.156,1.006]
Inaction Regret X Info		-1.140***		-1.092***
		[-1.890,-.390]		[-1.822,-.362]
(Experienced) Action Regret			-.556*	-.516*
			[-1.115,.003]	[-1.066,.034]
Action Regret X Info			1.101**	1.066**
			[.246,1.957]	[.224,1.909]
# Tasks encountered	.065**	.068**	.061**	.065**
	[.009,.121]	[.012,.123]	[.006,.116]	[.010,.120]
Risk Aversion	-.074	-.073	-.078*	-.078*
	[-.166,.018]	[-.166,.020]	[-.169,.013]	[-.171,.015]
Loss Aversion	.064	.069	.070	.077
	[-.072,.200]	[-.066,.204]	[-.062,.203]	[-.057,.210]
Constant	4.847***	4.656***	5.004***	4.810***
	[3.527,6.167]	[3.322,5.990]	[3.622,6.386]	[3.405,6.215]
Socio-demographic controls	Yes	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes	Yes
Observations	1719	1719	1719	1719

OLS Regressions.*** p <0.01, ***p* <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, representing the number of offers after which the participant stopped searching.Columns (1)-(4) display search behavior in tasks 2-10 and investigate the effect of regret experienced in the previous task. All columns include socio-demographic controls (gender, age, cognitive ability) and price sequence group fixed effects. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 2-10). *Risk Aversion* and *Loss Aversion* are defined as the number of safe choices, as described in Footnote 19.

A.4.4 Probit Regression: Optimality after Experienced Regret

We show that Table A.3 is robust to a probit specification in Columns (2)-(4).

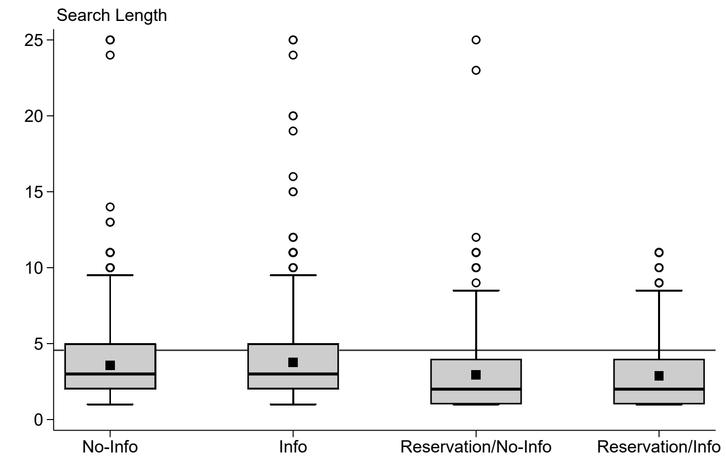
Table A.12: Probit Regression: Stopping the Search (Optimality)

	Forgone Profits	Optimal	Too few offers	Too many offers
	(1)	(2)	(3)	(4)
Treatments				
High-TP	.971	-.090**	.009	.075**
	[-.271,2.213]	[-.162,-.018]	[-.051,.069]	[.010,.139]
Info	.173	.016	.006	-.026
	[-1.261,1.607]	[-.060,.091]	[-.057,.069]	[-.100,.047]
High-TP X Info	-.606	.009	.005	-.009
	[-2.874,1.662]	[-.095,.113]	[-.080,.090]	[-.098,.080]
(Experienced) Inaction Regret	1.915**	-.125***	-.019	.111***
	[.399,3.430]	[-.216,-.034]	[-.090,.051]	[.050,.172]
Inaction Regret X Info	-1.505	.069	.028	-.062
	[-3.502,.492]	[-.049,.187]	[-.071,.126]	[-.141,.017]
(Experienced) Action Regret	-1.136	.080**	.007	-.087***
	[-2.669,.397]	[.005,.155]	[-.058,.073]	[-.146,-.028]
Action Regret X Info	3.526**	-.110**	.005	.110***
	[.851,6.200]	[-.219,-.001]	[-.085,.095]	[.029,.191]
# Tasks encountered	-.200**	.023***	-.031***	.009**
	[-.371,-.029]	[.015,.031]	[-.039,-.024]	[.002,.015]
Risk Aversion	-.160	.009	.003	-.010*
	[-.569,.250]	[-.005,.023]	[-.009,.015]	[-.020,.001]
Loss Aversion	-.167	-.010	-.007	.014
	[-.640,.307]	[-.030,.011]	[-.023,.009]	[-.004,.033]
Constant	6.334**			
	[.324,12.344]			
Socio-demographic controls	Yes	Yes	Yes	Yes
Price Sequence Group FE	Yes	Yes	Yes	Yes
Observations	1719	1719	1719	1719

*** $p<0.01$, **$p<0.05$, * $p<0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. Columns (1) shows an OLS regression, estimating the forgone profits compared to the ex-ante optimal benchmark. Column (2) estimates the likelihood that search behavior was optimal (compared to the ex-ante optimal benchmark) with a probit regression. The (binary) dependent variable takes the value 1 if the participant requested the optimal number of offers in the task and 0 otherwise. Column (3) shows the corresponding analysis with the dependent variable taking the value 1 if too few offers were requested and 0 otherwise. In Column (4), the dependent variable takes the value 1 if too many offers were requested and 0 otherwise. Columns (2)-(4) show marginal effects at the mean. All columns refer to search behavior in tasks 2-10. *(Experienced) Inaction Regret* is an indicator variable, taking a value of 1 if the participant experienced *inaction regret* in the previous task. *Inaction Regret X Info* is an indicator, taking a value of 1 if the participant experienced *inaction regret* in the previous task and was randomly assigned to treatments *Info*. *(Experienced) Action Regret* and *Action Regret X Info* are defined accordingly. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

A.5 Non-binding Reservation Prices

In our main experiment, we show that future price realizations do not alter search behavior when participants choose whether to buy the product or to continue the search after every offer. To study the robustness of our results, we introduce an additional pre-registered experiment in which participants repeatedly choose their reservation price before every offer in Section 1.6.3. Figure A.3 summarizes the main findings from this experiment. Most importantly, anticipating post-purchase information does not increase search length if participants repeatedly choose their reservation price before every offer (*Reservation/No-Info*: 2.93 vs. *Reservation/Info*: 2.88; $p = 0.719$, MWU). Interestingly, a direct comparison of both elicitation procedures shows that when participants make their choices through a reservation price, they request fewer offers. This holds both true with information (*Info:* 3.77 vs. *Reservation/Info:* 2.88; $p < 0.001$, MWU) and without information (*No-Info:* 3.56 vs. *Reservation/No-Info:* 2.93; $p = 0.001$, MWU) about future price realizations. Regression analyses (Table A.13) show that this effect is less pronounced in the first search task (Task 1).



Notes. The figure shows boxplots of search lengths across treatments in the additional experiment. The vertical line that indicates the optimal (ex-ante) threshold of a risk-neutral regret-free participant. The length of the whiskers is 1.5 times the interquartile range. The mean search length of each treatment is indicated by a solid square. The vertical line within the box corresponds to the median.

Figure A.3: Search Length across Information Structures and Elicitation Procedures (Tasks 1-10).

Finally, the replication of our baseline treatments (*Info* and *No-Info* without time pressure) show that search behavior is unaffected by the provision of post-purchase information. Average search lengths are similar (*Info:* 3.77 vs. *No-Info:* 3.56; $p = 0.418$, MWU), and payoffs closely aligned (*Info:* 25.20 vs. *No-Info:* 24.96; $p = 0.739$, MWU) across the two information structures. Regression analyses (Table A.13) corroborate these non-parametric results.

Table A.13: Search Length (Non-binding Reservation Prices)

	Number of offers					
	Task 1-10			Task 1		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments						
Reservation Price	-.627**	-.619***	-.553**	-.313	-.205	-.363
	[-1.104,-.150]	[-1.084,-.155]	[-.995,-.112]	[-1.059,.434]	[-.901,.491]	[-1.131,.404]
Info	.212	.213	.156	-.021	-.013	.166
	[-.296,.721]	[-.278,.705]	[-.345,.658]	[-.761,.719]	[-.756,.729]	[-.638,.970]
Reservation X Info	-.267	-.253	-.267	.000	-.016	-.053
	[-.929,.396]	[-.885,.378]	[-.838,.303]	[-1.054,1.054]	[-1.082,1.051]	[-1.073,.968]
# Tasks encountered	.070***	.070***	.070***			
	[.032,.107]	[.032,.107]	[.032,.107]			
Risk Aversion		.055	.072		.227***	.234***
		[-.033,.143]	[-.017,.160]		[.091,.364]	[.104,.365]
Loss Aversion		-.097	-.095		.053	.103
		[-.276,.083]	[-.266,.076]		[-.177,.284]	[-.130,.336]
Constant	3.175***	3.153***	3.343***	3.292***	1.198	1.026
	[2.807,3.543]	[2.019,4.286]	[2.133,4.553]	[2.750,3.833]	[-.734,3.130]	[-1.065,3.118]
Socio-demographic controls	No	Yes	Yes	No	Yes	Yes
Price Sequence Group FE	No	No	Yes	No	No	Yes
Observations	1920	1920	1920	192	192	192

OLS Regressions.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. Values in square brackets represent the 95% confidence intervals. The dependent variable is a count variable, which represents the number of offers after which the participant stopped searching. Columns (1)-(3) display search behavior across tasks 1-10, columns (4)-(6) for the first task. Columns (1) and (4) show the effect of the treatments. Columns (2) and (5) add socio-demographic controls (gender, age, cognitive ability) elicited after all search tasks; columns (3) and (6) additionally include price sequence group fixed effects. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points.

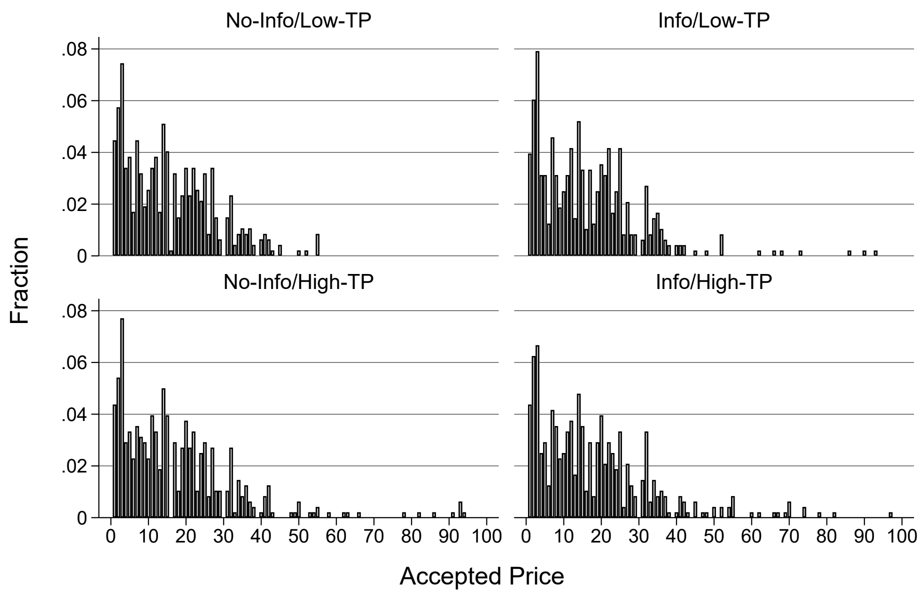
A.6 Search Heuristics

We shed more light on individual search behavior related to i) salient stopping prices, ii) bounce-heuristics, and iii) streak-based heuristics across treatments.

A.6.1 Salient Stopping Prices

First, we look at whether stopping behavior around salient reservation prices differs across experimental treatments. To do so, we define salient unfavorable prices as prices that always leads to a negative payoff irrespective of search costs and search length (i.e., prices larger than 50) and salient favorable prices as prices at or below 10. Overall, the probability to stop searching with all received price offers larger than 50 is very low (2.3 percent). In *Low-TP*, only 0.8 percent of stopping decisions happen with salient unfavorable prices (*Low-TP*/*No-Info*: 0.6 percent and *Low-TP*/*Info*: 1.0 percent). in *High-TP* this fraction amounts to 3.6 percent (*High-TP*/*No-Info*: 3.1 percent, *High-TP*/*Info*: 4.2 percent). Hence, with time pressure, salient unfavorable prices are more likely to be accepted ($p > 0.001$, MWU). Across information conditions, we do not find significant differences with respect to the acceptance of salient unfavorable prices (*No-Info*: 1.9 percent vs. *Info*: 2.6 percent; $p = 0.678$, MWU). Salient favorable prices are accepted with a much higher probability across all treatments (in 89 percent of the decisions) but mistakes are again more likely to occur with time pressure. Participants in *Low-TP* accept prices at or below 10 in 90.5 percent of the cases, participants in *High-TP* in 87.5 percent, with the difference being marginally statistically significant ($p < 0.060$). Again, we do not find strong differences across information conditions (*No-Info*: 87.3 percent vs. *Info*: 90.6 percent; $p = 0.254$, MWU).

An alternative way of studying whether particular salient prices influence stopping is to compare accepted prices across treatments. Figure A.4 shows histograms of accepted prices and highlights that the distribution is very similar across feedback (*No-Info* vs. *Info*; $p = 0.837$ Kolmogorov-Smirnov) and time pressure conditions (*Low-TP* vs. *High-TP*; $p = 0.388$ Kolmogorov-Smirnov test). Hence, we do not observe more frequent stopping at some salient cutoffs (e.g., 20/30/40) in particular treatments.



Notes. The figure shows the histograms of accepted prices across tasks 1-10 in each of the four treatment conditions.

Figure A.4: Accepted Prices across Treatments (Tasks 1-10).

Lastly, we expand our regression analysis on individual stopping behavior in Table A.14. In Column (1), we corroborate that participants in *High-TP* are somewhat less responsive to the current price. This is consistent with the finding that participants in *Low-TP* make fewer mistakes in the sense of accepting salient unfavorable prices or rejecting salient favorable prices.

A.6.2 Bounce Heuristics

"Bounce" heuristics describe individual search behaviors where the search was continued beyond the ultimately accepted price (e.g., a once-bounce heuristic could refer to "Have at least 2 searches and stop if a price quote larger than the previous quote is received", see Houser and Winter (2004) and Schunk and Winter (2009)). In the main part of the paper, we discussed the use of the recall option, which may reflect such bounce heuristics (i.e., "Stop at a price which is higher than the best price you have encountered so far."). We find that recall rates do not differ substantially across treatments. In *No-Info*, participants exercise the recall option in 18.8 percent of decisions, in *Info* in 17.9 ($p = 0.998$; MWU). In *High-TP*, rates are somewhat higher (20.5 percent) than in *Low-TP* with 16.2 percent ($p = 0.094$; MWU). Similarly, when focusing on other bounce heuristics, treatment differences are small. Analyzing the one-bounce heuristics following Houser and Winter (2004) and Schunk and Winter (2009), overall 10.9 percent of decisions are consistent with the one-bounce strategy: "Have at least 2 searches and stop if a price quote larger than the previous quote is received.", but we do not find treatment differences across feedback conditions (*No-Info*: 10.9 percent vs. *Info*: 10.8 percent; $p = 0.936$, MWU), and only small differences across time pressure conditions (*Low-TP*: 9.7 percent vs. *High-TP*: 12.1 percent; $p = 0.093$, MWU). We also analyze a modified one-bounce rule: "Have at least 2 searches and stop if a price quote larger than the previous quote less the search cost is received." Also here, we find no differences across feedback conditions (No-Info: 11.7 percent vs. Info: 11.4 percent; $p = 0.821$, MWU) and minor differences between *High-TP* and *Low-TP* (*Low-TP*: 10.1 vs. *High-TP*: 12.9; $p = 0.054$, MWU).

A.6.3 Streak-based Heuristics

In addition, we investigate how streaks in unfavorable past prices impact stopping behavior, akin to the idea of (losing) streak-based heuristics proposed in the literature (e.g., Houser & Winter, 2004; Schunk & Winter, 2009). Table A.14, Column (2) shows that, across treatments, participants are equally likely to stop after they encountered two times an unfavorable price in a row. Columns (3)-(6) confirm that this holds across treatments. Based on our analyses, we do not find convincing evidence that participants resort to different heuristics across treatments.

Table A.14: Probit Regression: Stopping the search (StreaK)

	1[Stopped Search]					
	Full Sample		No-Info	Info	Low-TP	High-TP
	(1)	(2)	(3)	(4)	(5)	(6)
Treatments						
High-TP	-.077*	-.007	-.010	.004		
	[-.158,.005]	[-.043,.028]	[-.042,.022]	[-.030,.039]		
Info	.040	.001			.001	.014
	[-.039,.119]	[-.032,.034]			[-.029,.032]	[-.023,.051]
High-TP X Info	.014	.012				
	[-.037,.065]	[-.039,.063]				
Price	-.009***	-.009***	-.009***	-.009***	-.009***	-.008***
	[-.011,-.008]	[-.010,-.008]	[-.010,-.007]	[-.010,-.008]	[-.010,-.008]	[-.009,-.007]
Price X High-TP	.002**					
	[.000,.004]					
Price X Info	-.001					
	[-.003,.001]					
Previous Two Prices[≥50]		-.004	-.007	-.002	-.014	.007
		[-.027,.018]	[-.037,.023]	[-.036,.032]	[-.043,.014]	[-.027,.041]
# Task encountered	-.005***	-.005***	-.005**	-.005**	-.004*	-.007***
	[-.008,-.002]	[-.009,-.002]	[-.010,-.001]	[-.010,-.000]	[-.008,.000]	[-.011,-.002]
Risk Aversion	.004	.004	-.004	.011**	.000	.005
	[-.003,.012]	[-.003,.012]	[-.013,.005]	[.000,.021]	[-.011,.012]	[-.004,.013]
Loss Aversion	.000	.001	.006	-.010	.001	-.000
	[-.010,.011]	[-.009,.011]	[-.006,.018]	[-.026,.006]	[-.018,.020]	[-.012,.011]
# of choices	7226	7226	3643	3583	3591	3635
Price FE	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-sq	.3	.3	.28	.33	.35	.26

Probit Regression.*** p <0.01, **p <0.05, * p <0.1. Standard errors clustered at the individual level. The table shows marginal effects at the mean from a probit regression. Columns (1) & (2) display search behavior across all treatments, columns (3) & (4) in the respective feedback environment. *Price* is the price of the current offer [1,100] the participant faces. *Previous Two Prices[≥50]* is an indicator variable taking the value 1 if the previous two prices were ≥50. *# Tasks encountered* is a count variable, indicating the number of the current task (Task 1-10). *Risk Aversion* and *Loss Aversion* are defined as switching points, as described in Footnote 19.

A.7 Instructions

A.7.1 Main Experiment

Appendix A.7.1 includes the translated instructions of the main experiment (from German). The participants received the instructions for the experiment in print. Additional short instructions and control questions were later displayed on the computer screen. Treatment specific parts are shown in *italics* and the corresponding treatment clearly indicated.

Welcome to the experiment and thank you for your participation!

Please do not speak from now on with any other participant

General Procedures

In this experiment, we study economic decision-making. You can earn money by participating. The money you earn will be paid to you privately and in cash after the experiment. The experiment lasts for around 60 minutes and consists of multiple parts (the exact number of parts is unknown to all participants). At the beginning of every part, you receive detailed instructions. If you have questions after reading the instructions or during the experiment, please raise your hand or press the red button on your keyboard. One of the experimenters will then come to you and answer your question(s) privately.

During the experiment, you and the other participants will be asked to make decisions. These can affect the payoffs for you, and potentially for other participants. How your decisions relate to the payoffs will be explained in more detail in the instructions (or later on the screen).

Important: Depending on the decision, you will see an expiring clock at two different places on the screen. If you see the clock with the tag "Remaining time" in the center of the screen it indicates how much time you have for the decision. Further information will be provided in the instructions.
During other decisions, you will see a (small) expiring clock at the right-upper part

of the screen. This time only gives you an indication, how long the current decision should take. You can also take more time if you need it. Entering a decision is also possible before time expires.

Anonymity

The analysis of the experiment is anonymous; that is, we will never link your name with the data generated in the experiment. At the end of the experiment, you will be asked to sign a receipt to confirm the payments you received. This receipt will only be used for accounting purposes. No further personal data will be passed on.

Tools

You find a pen at your desk. Please leave the pen and the instructions on the table after the experiment.

Payment

In addition to the income that you earn during the experiment, you will receive 6 € for showing up on time. During the experiment, we do not talk about Euro, but about Taler. We convert the Taler into Euros at the end of the experiment and pay those in addition to the 6 € for your punctual appearance in cash.

Procedure

This experiment consists of **multiple decisions on the purchase of a fictitious product**. In the following, the rules that determine the payoff from your decisions, are explained in detail. At the end of the experiment, one of the buying decisions will be randomly chosen and you receive the corresponding payoff. Every purchase decision is equally likely to be randomly chosen.

After the purchase decisions, you can earn additional money through correct assessments and further decisions.

Following this, we will ask you to respond to a few questions conscientiously. After that, the experiment ends. You will then receive the money that you earned through your decisions, as well as 6€ in cash for your punctual appearance.

Exchange rate in the purchase decisions

In some parts of the experiment, we do not task about Euros, instead we refer to Taler. These will be converted into Euros at the end of the experiment. Please note the following exchange rate:

100 Taler = 12 €

Your task

The experiment has several tasks. In every task, the objective is to obtain as many Taler as possible through the purchase of a fictitious product. In general, a task proceeds as follows.

In every task, the number of Taler you receive from a purchase decision is calculated as the difference between the value of the product and the costs that you incur through making the purchase.

Taler from the purchase decision = Value of the product – Price – Cost for price offers

Value of the product

The **product is worth 50 Taler for you.**

When you buy the product, **you receive 50 Taler**. At the same time, you have to pay a price for the purchase of the product.

Price of the product and cost for the price offers

The computer offers the product to you by displaying a purchase price, at which you can buy the product. You can then decide whether you want to request another offer in the form of a new purchase price or whether you want to buy the product for the lowest purchase price offered so far. You can request as many offers as you want (as long as there is a possibility to achieve a positive payoff under any search cost). However, every offer you request is associated with a cost for you:

Every offer you request costs a fixed amount of Taler.

In the following, **these costs will be called search costs**. The search cost can vary across tasks. You will know the exact cost level before each purchase decision.
You **can always buy the product at the lowest standing offer** (even if you have requested additional offers that might have been higher). Therefore, amount of Taler you receive from a purchase decision is

50 – (lowest price received) – search cost*(number of offers you requested).

Accordingly, the amount of Taler you receive is higher when the price at which you purchase the product is lower. The amount of Taler decreases by the amount of search cost with every offer you request. (For the first, automatically displayed offer, you do **not** incur any costs.)

Time for the decision

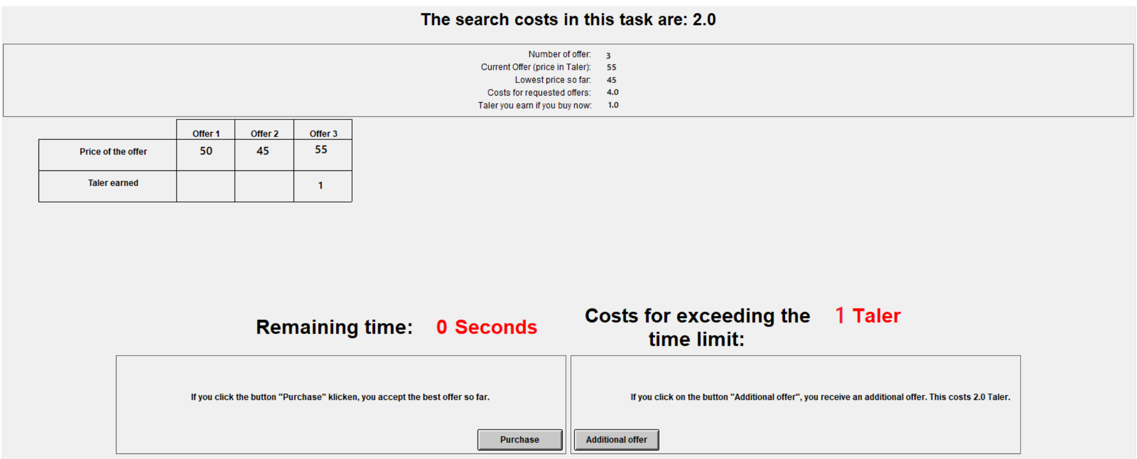
You only have limited time to make your decision. After every offer you have *60 seconds [Low-TP/No-Info and Low-TP/Info]/ 4 seconds [High-TP/No-Info and High-TP/Info]* to decide whether you want to buy for the best price observed so far or whether you want to request another offer. If you neither decide to buy the product nor request an additional offer, we will deduct 1 Taler from your payoff in this task. Afterward, you have an additional *60 seconds [Low-TP/No-Info and Low-TP/Info]/ 4 seconds [High-TP/No-Info and High-TP/Info]* to make the decision (purchasing vs. requesting another offer). If you do not decide within that time once again, you will be again deducted 1 Taler in this task. This procedure is repeated until you make a decision.

Information on the offers of the computer

The price offers of the computer are integers and can take the values 1, 2, 3... to 100 Taler. The computer draws each price independently and randomly with the same probability of 1% (draws with replacement). You can imagine the procedure like this: an urn contains 100 balls, which are numbered from 1 to 100. At each offer, the computer draws one of those balls, displays the number on the ball as a price offer, and puts the ball back into the urn, such that each ball in the next draw will be again drawn with a probability of 1%.

On-screen procedure

To illustrate the decision screen, below you can see an example of a task, where—in addition to the first offer of the computer (price of 50)—two more offers were requested:



In the upper part, you see the search cost for this task. Below you see how many offers are already displayed, as well as which offer is the current offer and which is the best one. Additionally, you see the costs that have to be paid for the offers requested so far.

In the lower part you make your purchase decision. To accept the best offer so far, you click on the button: “Buy”. To request another offer and incur the above-displayed search cost, you click on the button: “Additional offer”.

In the central part, you see an overview of the offers received so far, as well as your current payoff for the task if you click “Buy.”

In the displayed example, the first offer was equal to 50 Taler. Because the product is worth 50 Taler, buying the product at this price would have resulted in a payoff of 0 Taler in this task. In the example, we assumed, that another offer was requested at the (search) cost of 2 Taler.

The second price offered to you, was 45 Taler in the example. Deciding to buy at this offer would have led to receiving the product at the lowest price so far observed (i.e., 45 Taler). Hence, your payoffs would have been determined as follows:

Received Taler	=	value of the product	–	lowest price	–	search cost (2 Taler for each requested offer)
	=	50	–	45	–	2 = **3**

In the example, we assumed that another offer at the cost of 2 Taler was requested. This time, the randomly drawn price was 55 Taler. If you decided to purchase the product at this point within the remaining time, then you would receive 1 Taler for this task (as you can always purchase the product for the lowest price seen so far):

Received Taler	=	value of the product	–	lowest price	–	search cost (2 Taler for each requested offer)
	=	50	–	45	–	2*2 = **1**

If you instead requested another offer, then you would incur the cost of 2 Taler again and the computer would display an additional randomly drawn price.

Beneath the offers seen so far, you see the “Remaining Time” for the decision. This shows how much time you have remaining to decide between “Buy” and “Additional offer”. On the right-hand side, you see how many Taler were already deducted from your payoff due to exceeding the time limit in this task.

In the example, we assumed that the decision time has just expired, such that an additional cost of 1 Taler through exceeding the time limit has to be paid. After the expiration of the decision time, the “Remaining time” further runs down. Should you decide to buy the product after offer 3 in the next 60 seconds, you receive 0 Taler in this task. Should you request another offer within this time, then you pay the search cost of 2 Taler and the computer displays an additional randomly drawn price. Should you neither buy the product nor request another offer within the next 60 seconds, you incur a cost of 1 Taler again. This procedure is repeated until you make a decision.

Note

In every task it is possible, that you receive a negative payoff. If this task is drawn as payoff relevant, this loss will be offset by your payoff from the other parts of the experiment.

Procedure

After every purchase decision, you will see all the offers until your purchase decision once again. *Furthermore, you see additional offers, which would have been displayed to you later, if you had not made a purchase decision at that point. This means, you will see whether requesting one or multiple additional offers would have yielded more (or less) Taler. [only in Low-TP/Info and High-TP/Info]*
To conclude the task, please type in the number of the offer, at which you would have received the highest payoff.

After the purchase decisions, you will be additionally asked for assessments of your own behavior and you will be asked to make additional decisions, with which you can earn or lose money. At the end of the experiment, you see your payoff on a separate screen. You will also be shown, which of the purchase decisions has been randomly drawn to be relevant for your payoff.

Comprehension questions

To verify your understanding of the task and the payoff scheme, you will be confronted with some control questions before the purchase decisions start. The first purchase decision starts when all participants have answered the questions correctly. Important: Your answers to the comprehension questions do not affect your payoff.

Additional On-screen Instructions

Expected Performance; rank in own treatment

You made several purchase decisions in the first part of the experiment.
Please think back to the **first 10 purchase decisions**, where you could decide after

each offer whether to accept it or not.
5 other people have seen the same price sequences as you in this part.
Below we ask you to rate how successful you were in this part compared to the other people.
For this, we have calculated the **average payout of all 10 rounds**.
Below we ask you to rate how successful you were in this part compared to the other people.
For a correct estimation, you will receive 2 EUR. Otherwise, you will receive 0 EUR.
Estimate your rank based on the **average payout**:

◦ 1 ◦ 2 ◦ 3 ◦ 4 ◦ 5 ◦ 6

Expected Performance; rank in opposite time-pressure treatment

Please think back again to the **first 10 purchase decisions**, where you could decide after each offer whether to accept it or not.
We ask you again to compare yourself with 5 other people.
These have also seen the same price sequences.
However, these participants each had ***60 seconds*** *[High-TP/No-Info and High-TP/Info]*/ ***4 seconds*** *[Low-TP/No-Info and Low-TP/Info]* to make a decision.
As a reminder, you had *4 seconds [High-TP/No-Info and High-TP/Info]/ 60 seconds [Low-TP/No-Info and Low-TP/Info].*
Below we ask you to rate how successful you were in this part compared to the other people.
Unlike the previous decision, this question is hypothetical and you will not receive a payout based on your answer.
Nevertheless, your answer to this question is of great interest.
Estimate your rank based on the **average payout**:

◦ 1 ◦ 2 ◦ 3 ◦ 4 ◦ 5 ◦ 6

Loss attitudes [Task A] (Gächter, Johnson, & Herrmann, 2022)

Task A consists of 6 decisions where you can accept up to 6 offers.
The offers consist of a lottery through which you can lose or win money. You have to decide for each of the 6 offers whether to accept it or not.
For each accepted offer, the computer loses or wins an amount of money.
At the end of the experiment, your decision is implemented for one of the 6 offers. The computer randomly selects (with equal probability) which offer will be implemented.

Decide for each offer whether you want to accept it.

1	With 50% probability you lose 2 euros; with 50% probability you win 6 euros.	∘ accept ∘ reject
2	With 50% probability you lose 3 euros; with 50% probability you win 6 euros.	∘ accept ∘ reject
3	With 50% probability you lose 4 euros; with 50% probability you win 6 euros.	∘ accept ∘ reject
4	With 50% probability you lose 5 euros; with 50% probability you win 6 euros.	∘ accept ∘ reject
5	With 50% probability you lose 6 euros; with 50% probability you win 6 euros.	∘ accept ∘ reject
6	With 50% probability you lose 7 euros; with 50% probability you win 6 euros.	∘ accept ∘ reject

Risk attitudes [Task B] (Holt & Laury, 2002)

Task B consists of 10 decisions, each of which allows you to choose between 2 offers. The offers consist of a lottery through which you win money. You must choose lottery X or Y for each of the 10 choices.
For each lottery you choose, the computer will draw the amount of money you win.
At the end of the experiment, one of the 10 decisions is implemented. The computer randomly selects (with equal probability) which decision will be implemented.

Decide in each case whether you want to accept X or Y.

	Option X	Option Y	
1	With 10% probability you win 2.00 Euro; with 90% probability you win 1.60 Euro.	With 10% probability you win 3.85 Euro; with 90% probability you win 0.10 Euro.	∘ X ∘ Y
2	With 20% probability you win 2.00 Euro; with 80% probability you win 1.60 Euro.	With 20% probability you win 3.85 Euro; with 80% probability you win 0.10 Euro.	∘ X ∘ Y
3	With 30% probability you win 2.00 Euro; with 70% probability you win 1.60 Euro.	With 30% probability you win 3.85 Euro; with 70% probability you win 0.10 Euro.	∘ X ∘ Y
4	With 40% probability you win 2.00 Euro; with 60% probability you win 1.60 Euro.	With 40% probability you win 3.85 Euro; with 60% probability you win 0.10 Euro.	∘ X ∘ Y
5	With 50% probability you win 2.00 Euro; with 50% probability you win 1.60 Euro.	With 50% probability you win 3.85 Euro; with 50% probability you win 0.10 Euro.	∘ X ∘ Y
6	With 60% probability you win 2.00 Euro; with 40% probability you win 1.60 Euro.	With 60% probability you win 3.85 Euro; with 40% probability you win 0.10 Euro.	∘ X ∘ Y
7	With 70% probability you win 2.00 Euro; with 30% probability you win 1.60 Euro.	With 70% probability you win 3.85 Euro; with 30% probability you win 0.10 Euro.	∘ X ∘ Y
8	With 80% probability you win 2.00 Euro; with 20% probability you win 1.60 Euro.	With 80% probability you win 3.85 Euro; with 20% probability you win 0.10 Euro.	∘ X ∘ Y
9	With 90% probability you win 2.00 Euro; with 10% probability you win 1.60 Euro.	With 90% probability you win 3.85 Euro; with 10% probability you win 0.10 Euro.	∘ X ∘ Y
10	With 100% probability you win 2.00 Euro; with 0% probability you win 1.60 Euro.	With 100% probability you win 3.85 Euro; with 0% probability you win 0.10 Euro.	∘ X ∘ Y

Socio-demographics

Please provide the following statistical information.

- Gender [male; female]

- Age [integer]
- Field of study (faculty/major)
  - 1=Humanities
  - 2=Engineering
  - 3=Medicine
  - 4=Natural Science
  - 5=Law
  - 6=Economics
  - 7=Social Science
  - 8=Other
- What is your high school graduation grade in mathematics? [integer; 1-6]
- What language(s) is (are) your native language(s)? [string]
- How many times have you participated in an economic laboratory study (including outside of this laboratory)? [integer]
- How many participants from the experiment do you know personally? [integer]
- If there is anything else you would like to tell us regarding the experiment, please enter it here: [string]

A.7.2 Additional Experiment: Non-binding Reservation Price

Appendix A.7.2 includes the translated instructions for the treatments with a repeated reservation price elicitation (*Reservation/Info* and *Reservation/No-Info*) of the additional experiment (from German).

Welcome to the experiment and thank you for your participation!

Please do not speak from now on with any other participant

General Procedures

In this experiment, we study economic decision-making. You can earn money by participating. The money you earn will be paid to you privately and in cash after the experiment. The experiment lasts for around 60 minutes and consists of multiple parts (the exact number of parts is unknown to all participants). At the beginning of every part, you receive detailed instructions. If you have questions after reading the instructions or during the experiment, please raise your hand or press the red button on your keyboard. One of the experimenters will then come to you and answer your question(s) privately.

During the experiment, you and the other participants will be asked to make decisions. These can affect the payoffs for you, and potentially for other participants. How your decisions relate to the payoffs will be explained in more detail in the instructions (or later on the screen).

Important: Depending on the decision, you will see an expiring clock at two different places on the screen. If you see the clock with the tag "Remaining time" in the center of the screen it indicates how much time you have for the decision. Further information will be provided in the instructions.
During other decisions, you will see a (small) expiring clock at the right-upper part of the screen. This time only gives you an indication, how long the current decision should take. You can also take more time if you need it. Entering a decision is also possible before time expires.

Anonymity

The analysis of the experiment is anonymous; that is, we will never link your name with the data generated in the experiment. At the end of the experiment, you will be asked to sign a receipt to confirm the payments you received. This receipt will only be used for accounting purposes. No further personal data will be passed on.

Tools

You find a pen at your desk. Please leave the pen and the instructions on the table after the experiment.

Payment

In addition to the income that you earn during the experiment, you will receive 6 € for showing up on time. During the experiment, we do not talk about Euro, but about Taler. We convert the Taler into Euros at the end of the experiment and pay those in addition to the 6 € for your punctual appearance in cash.

Procedure

This experiment consists of **multiple decisions on the purchase of a fictitious prod-**

uct. In the following, the rules that determine the payoff from your decisions, are explained in detail. At the end of the experiment, one of the buying decisions will be randomly chosen and you receive the corresponding payoff. Every purchase decision is equally likely to be randomly chosen.
After the purchase decisions, you can earn additional money through correct assessments and further decisions.
Following this, we will ask you to respond to a few questions conscientiously. After that, the experiment ends. You will then receive the money that you earned through your decisions, as well as 6€ in cash for your punctual appearance.

Exchange rate in the purchase decisions

In some parts of the experiment, we do not task about Euros, instead we refer to Taler. These will be converted into Euros at the end of the experiment. Please note the following exchange rate:

100 Taler = 12 €

Your task

The experiment has several tasks. In every task, the objective is to obtain as many Taler as possible through the purchase of a fictitious product. In general, a task proceeds as follows.

In every task, the number of Taler you receive from a purchase decision is calculated as the difference between the value of the product and the costs that you incur through making the purchase.

Taler from the purchase decision = Value of the product – Price – Cost for price offers

Value of the product

The **product is worth 50 Taler for you.**

When you buy the product, **you receive 50 Taler**. At the same time, you have to pay a price for the purchase of the product.

Price of the product and cost for the price offers

The computer offers you the product. It makes you offers in the form of purchase prices at which you can buy the product. In the process, the computer makes one offer after another. Offers made remain valid, so you can always purchase the product at the lowest purchase price offered so far. However, every offer is associated with a cost for you. This means that you can receive as many offers as you want (as long as there is a possibility of achieving a positive payout amount in one round), but you pay for each offer:

Every offer you request costs a fixed amount of Taler.

In the following, **these costs will be called search costs**. The search cost can vary across tasks. You will know the exact cost level before each purchase decision.
You **can always buy the product at the lowest standing offer** (even if you have requested additional offers that might have been higher). Therefore, amount of Taler you receive from a purchase decision is

50 – (lowest price received) – search cost*(number of offers you requested).

Accordingly, the amount of Taler you receive is higher when the price at which you purchase the product is lower. The amount of Taler decreases by the amount of search cost with every offer you request. (For the first, automatically displayed offer, you do **not** incur any costs.)

Your purchase decision

Before each offer you receive from the computer, you specify your *maximum purchase price*. Your *maximum purchase price* determines the price up to which you would buy the product. If the computer's next offer is lower than (or equal to) your *maximum purchase price*, you buy the product at the offered purchase price. If the offer is higher than your *maximum purchase price*, you will not buy the product. In this case, you will be asked again to enter a *maximum purchase price*. This can be different from your last entry, but it does not have to be. After that, you will receive another offer (for which you will pay the search cost displayed on the screen).

Time for the decision

You only have limited time to make your decision. After every offer you have 60 seconds to decide whether you want to buy for the best price observed so far or whether you want to request another offer. If you neither decide to buy the product nor request an additional offer, we will deduct 1 Taler from your payoff in this task. Afterward, you have an additional 60 seconds to make the decision (purchasing vs.

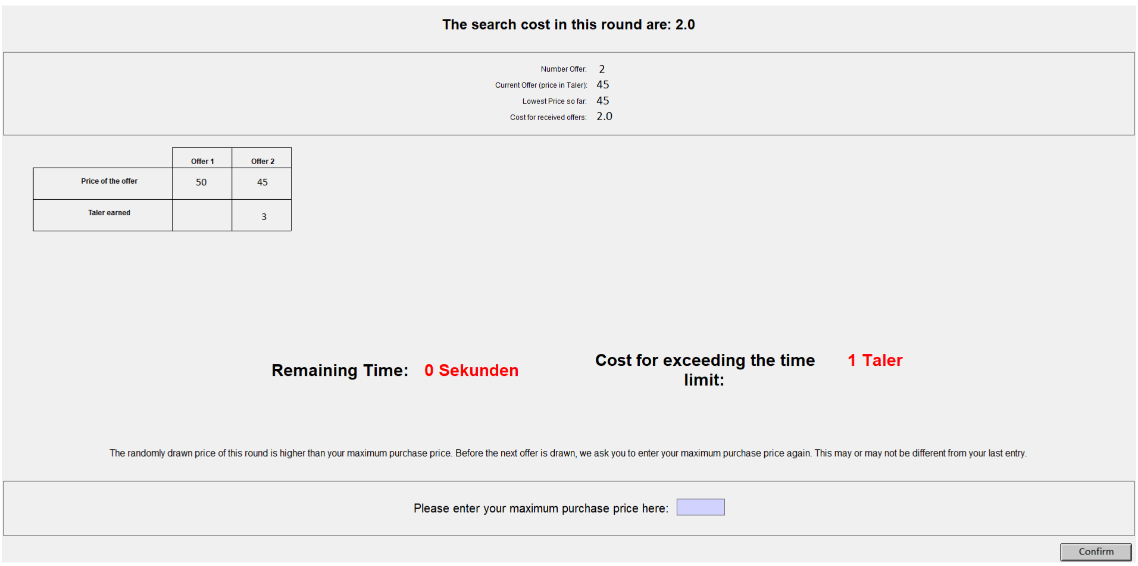
requesting another offer). If you do not decide within that time once again, you will be again deducted 1 Taler in this task. This procedure is repeated until you make a decision.

Information on the offers of the computer

The price offers of the computer are integers and can take the values 1, 2, 3... to 100 Taler. The computer draws each price independently and randomly with the same probability of 1% (draws with replacement). You can imagine the procedure like this: an urn contains 100 balls, which are numbered from 1 to 100. At each offer, the computer draws one of those balls, displays the number on the ball as a price offer, and puts the ball back into the urn, such that each ball in the next draw will be again drawn with a probability of 1%.

On-screen procedure

To illustrate the decision screen, below you can see an example of a task, where—in addition to the first offer of the computer (price of 50)—another offer has already been made by the computer:



In the upper part, you see the search cost for this task. Below you see how many offers are already displayed, as well as which offer is the current offer and which is the best one. Additionally, you see the costs that have to be paid for the offers received so far.

In the lower part you make your decision by entering your *maximum purchase price* in the free field. This can be between 1 and 100 Taler. To confirm it and receive the next offer, click "Confirm".

In the central part, you see an overview of the offers received so far, as well as the payoff you would have received if you had purchased at the current lowest price.

In the displayed example, the first offer was equal to 50 Taler. Because the product is worth 50 Taler, buying the product at this price would have resulted in a payoff of 0 Taler in this task. In the example, we assumed that your first *maximum purchase price* was lower than 50 Taler and therefore you did not buy the product. Instead, you were asked again for your *maximum purchase price*, entered it, and then received another offer at the cost of 2 Taler.

The second price offered to you, was 45 Taler in the example. If you had specified a *maximum purchase price* of 45 Taler or higher after the first offer, you would have received the product for the lowest price so far: 45 Taler. Hence, the achieved Taler would have been determined as follows:

Received Taler = value of the product – lowest price – search cost (2 Taler for each requested offer)						
=	50	–	45	–	2	= **3**

In the example we assumed that your *maximum purchase price* after the first offer was lower than 45. Therefore, you now enter your maximum purchase price again and then receive another offer at a cost of 2 Taler. This is the current situation that you see on the screenshot.

If you now enter a *maximum purchase price* of at least 45 Taler within the remaining time, you will receive the product regardless of the next offer (as you can always buy the product at the lowest price offered so far). Therefore, if the next price is higher than 45 Taler (for example, price = 55 Taler), you received 1 Taler for the purchase in this round:

Received Taler = value of the product – lowest price – search cost (2 Taler for each requested offer)						
=	50	–	45	–	2*2	= **1**

If instead you enter a *maximum purchase price* of less than 45 Taler and the next randomly drawn price is 55 Taler, you would not buy the product and then be asked for your *maximum purchase price*. By entering it again, you would incur additional cost of 2 Taler, and the computer would show you another randomly drawn price.

If offer 3 is less than 45 Taler you buy the product, provided that your *maximum purchase price* after the second offer was greater than this.

Beneath the offers seen so far, you see the "Remaining Time" for the decision. This shows how much time you have remaining to decide about your *maximum purchase price*. On the right-hand side, you see how many Taler were already deducted from your payoff due to exceeding the time limit in this task.

In the example, we assumed that the decision time has just expired, such that an additional cost of 1 Taler through exceeding the time limit has to be paid. After the expiration of the decision time, the "Remaining time" further runs down. If you then decide to make an entry within the next 60 seconds and it leads to a purchase at the price of 45 Taler, you will receive 0 Taler in this round (since the product is worth 50 Taler and you incurred a total search cost of 4 Taler, plus 1 Taler for exceeding the time limit). Should not make an entry again within 60 seconds, you incur a cost of 1 Taler again. This procedure is repeated until you make a decision.

Note

In every task it is possible, that you receive a negative payoff. If this task is drawn as payoff relevant, this loss will be offset by your payoff from the other parts of the experiment.

Procedure

After every purchase decision, you will see all the offers until your purchase decision once again. *Furthermore, you see additional offers, which would have been displayed to you later, if you had not made a purchase decision at that point. This means, you will see whether requesting an or multiple additional offers would have yielded more (or less) Taler. [only in Reservation/Info]*
To conclude the task, please type in the number of the offer, at which you would have received the highest payoff.

After the purchase decisions, you will be additionally asked for assessments of your own behavior and you will be asked to make additional decisions, with which you can earn or lose money. At the end of the experiment, you see your payoff on a separate screen. You will also be shown, which of the purchase decisions has been randomly drawn to be relevant for your payoff.

Comprehension questions

To verify your understanding of the task and the payoff scheme, you will be confronted with some control questions before the purchase decisions start. The first purchase decision starts when all participants have answered the questions correctly. Important: Your answers to the comprehension questions do not affect your payoff.

B

Appendix to Chapter 2

Everyone Likes to be Liked

B.1 Preregistered Analyses

B.1.1 Result 2: Regression Analysis

Table B.1: Preference Adjustments across Treatments

	1[Preference Adjustment]	
	(1)	(2)
Info	.165***	.158***
	[.095,.234]	[.089,.227]
Loss Aversion		-.020
		[-.051,.012]
Cognitive Ability (Raven's')		-.007
		[-.037,.022]
Male		-.002
		[-.073,.070]
Observations	575	575

Notes. Logit Regressions. *** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. The table shows marginal effects at the mean from a logit regression where the dependent variable is an indicator for whether someone changed their preferences.

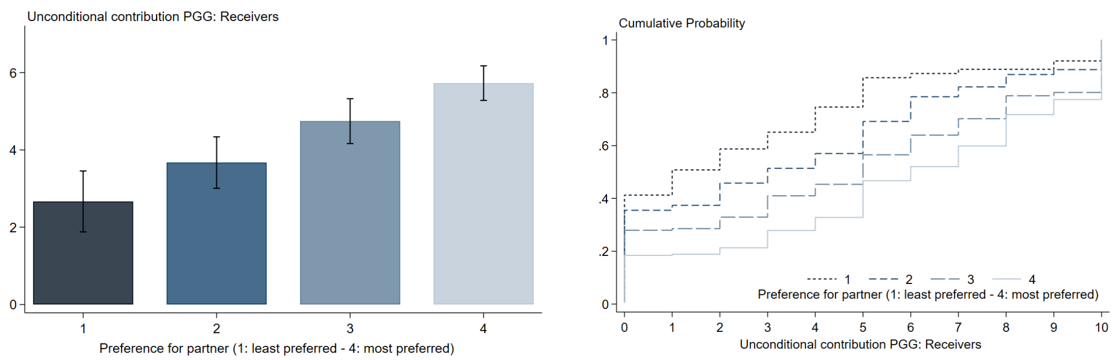
B.1.2 Result 3: Regression Analysis

Table B.2: Consistency of Preference Adjustments with Reciprocal Preferences

	1[Consistent Preference Adjustment]	
	(1)	(2)
Info	.152***	.150***
	[.108,.195]	[.107,.193]
Loss Aversion		-.010
		[-.024,.005]
Cognitive Ability (Raven's')		-.000
		[-.016,.016]
Male		-.021
		[-.060,.018]
Observations	575	575

Notes. Logit Regressions. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. This table shows marginal effects at the mean from logit regressions where the dependent variable is an indicator for whether someone changed their preferences consistent with having reciprocal preferences. *Info* is an indicator, taking the value of one if the participant was randomly assigned to the treatment *Info*. *Loss aversion* and *Cognitive ability* are calculated as detailed in Footnote 16, *Male* is an indicator taking the value of 1 if a participant indicated to identify as male.

B.1.3 Unconditional Contributions of Receivers



A) Averages by *Preference for partner*

B) Distributions by *Preference for partner*

Notes. This figure displays the unconditional contributions of receivers by their preferences for the matched proposer. *Preference for partner (1-4)* takes the value of four if the matched proposer was the first choice of the receiver, three if the matched receiver was the second choice, and so on. Panel A shows averages, Panel B the cumulative distribution functions.

Figure B.1: Unconditional PGG Contributions: Receiver

Table B.3: Unconditional PGG Contributions of Receivers

	Unconditional PGG Contribution (0-10)	
	(1)	(2)
Preference for partner (1-4)	1.023***	.960***
	[.719,1.328]	[.666,1.253]
Round		-.216***
		[-.345,-.087]
Loss Aversion		-.480*
		[-.987,.027]
Cognitive Ability (Raven's')		.315
		[-.090,.720]
Male		-.805
		[-2.207,.597]
Observations	575	575

Notes. OLS Regressions. *** p <0.01, ** p <0.05, * p <0.1. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. *Preference for partner (1-4)* takes the value of four if the matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. *Round* is a count variable, indicating the number of the current round (Round 1-5). *Loss aversion* and *Cognitive ability* are calculated as detailed in Footnote 16, *Male* is an indicator taking the value of 1 if a participant indicated to identify as male.

B.1.4 Mechanisms across Treatments

In Sections 2.4.2-2.4.4, we compared proposers ranked favorably to proposers ranked less favorably by their partner within *Info*. To corroborate these results and to substantiate that they are specific to the information environment in *Info*, we now analyze the effect of being ranked favorable across both information conditions. We compare beliefs and contributions in the situation in which proposers knew their partner's preference (*Info*) to that in which the proposers did not know it (*No-Info*). Hence, in a type of Placebo test, we estimate the effect of *knowing* the rank on contributions and beliefs while holding the actual rank received by the partner constant across treatments. Table B.4 shows that none of our variables of interest is significant in *No-Info*.

Table B.4: PGG Behavior of Proposers in *Info* and *No-Info*

	Belief Partner Contribution	Unconditional Contribution	Avg. Conditional Contribution
	(1)	(2)	(3)
Preference for partner (1-4)	.246	.129	.035
	[-.059,.551]	[-.172,.430]	[-.128,.198]
Partner's preference (1-4)	.106	-.068	-.053
	[-.265,.477]	[-.499,.363]	[-.339,.233]
Partner's Preference X Info	1.208***	.862***	.444**
	[.600,1.815]	[.248,1.475]	[.062,.826]
Info	-2.566**	-1.609	-.763
	[-4.704,-.429]	[-3.826,.609]	[-2.123,.596]
Round	-.021	-.254***	-.169***
	[-.141,.099]	[-.359,-.149]	[-.238,-.100]
Loss Aversion	-.609**	-.570**	.300*
	[-1.112,-.106]	[-1.130,-.009]	[-.041,.641]
Cognitive Ability (Raven's)	.272	.180	-.196
	[-.133,.678]	[-.315,.675]	[-.465,.073]
Male	-.928*	-.884	-.325
	[-1.974,.117]	[-2.155,.387]	[-1.118,.468]
Observations	575	575	575

Notes. OLS Regressions. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Standard errors clustered at the individual level. The values in square brackets represent the 95% confidence intervals. *Partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their matched partner, three if the participant was the second most preferred choice, and so on. *Preference for partner (1-4)* takes the value of four if the matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. The interaction term *Partner's Preference X Info* takes the value of zero for observation in *No-Info*, and the value of *Partner's Preference X (1-4)* in *Info*. *Info* is an indicator, taking the value of one if the participant was randomly assigned to the treatment *Info*. *Round* is a count variable, indicating the number of the current round (Round 1-5). *Loss aversion* and *Cognitive ability* are calculated as detailed in Footnote 16, *Male* is an indicator taking the value of 1 if a participant indicated to identify as male.

B.2 Theoretical Framework

B.2.1 Proof of Proposition 1

Proof. By Assumptions 1 and 2, an increase in l_r increases a_p and c_r. We use the Implicit Function Theorem to prove that an increase in a_p or c_r both weakly increases

c_p. Hence, the increase in l_r must weakly increase c_p. We start with Equation (2.2) derived in Section 2.4.1, which shows the condition that maximizes the adjusted utility of a proposer, assuming an interior solution.

$$F(c_p; a_p, c_r) = \frac{\partial v_p}{\partial c_p} = \underbrace{\frac{\partial u_p}{\partial c_p}}_{<0} + \underbrace{a_p \cdot \frac{\partial u_r}{\partial c_p}}_{>0} = 0$$

We can make statements about the first and second partial derivatives of the twice differentiable concave direct utility functions ($u_{p,r}$). Higher contributions by the proposer c_p increase the monetary outcome of a receiver and decrease the monetary outcome of a proposer. This means that a higher contribution by the proposer (c_p) has a negative effect on the proposer's direct utility, while it positively affects the receiver's direct utility ($\frac{\partial u_p}{\partial c_p} < 0, \frac{\partial u_r}{\partial c_p} > 0$). The second partial derivatives, $\frac{\partial^2 u_p}{\partial {c_p}^2} < 0$ and $\frac{\partial^2 u_r}{\partial {c_p}^2} < 0$, are both negative. The positive marginal utility of more money decreases for the receiver. The negative marginal utility of losing money increases with less money for the proposer. The mixed partial derivatives are both positive ($\frac{\partial^2 u_p}{\partial c_p \partial c_r} > 0$ and $\frac{\partial^2 u_r}{\partial c_p \partial c_r} > 0$). For higher contributions of the other player, the negative marginal utility of contributing to the PGG is smaller, because the income is higher. This is true for the proposer and the receiver.

We use the Implicit Function Theorem to show how a change of a_p and c_r affects c_r. Proof that the optimal contribution c_p increases with a higher level of altruism ($\frac{\partial c_p}{\partial a_p} > 0$):

$$\frac{\partial c_p}{\partial a_p} = -\frac{\frac{\partial F}{\partial a_p}}{\frac{\partial F}{\partial c_p}} = -\frac{\overbrace{\frac{\partial u_r}{\partial c_p}}^{>0}}{\underbrace{\frac{\partial^2 u_p}{\partial {c_p}^2}}_{<0} + \underbrace{a_p \frac{\partial^2 u_r}{\partial {c_p}^2}}_{<0}} > 0$$

Proof that the optimal contribution c_p increases with a higher contribution of the receiver ($\frac{\partial c_p}{\partial c_r} > 0$):

$$\frac{\partial c_p}{\partial c_r} = -\frac{\dfrac{\partial F}{\partial c_r}}{\dfrac{\partial F}{\partial c_p}} = -\frac{\overbrace{\dfrac{\partial^2 u_p}{\partial c_p \partial c_r}}^{>0} + a_p \overbrace{\dfrac{\partial^2 u_r}{\partial c_p \partial c_r}}^{>0}}{\underbrace{\dfrac{\partial^2 u_p}{\partial {c_p}^2}}_{<0} + a_p \underbrace{\dfrac{\partial^2 u_r}{\partial {c_p}^2}}_{<0}} > 0$$

The equations above show that the denominator $\partial F / \partial c_p$ is always smaller than 0. Therefore, the necessary condition for the Implicit Function Theorem holds that the denominator is never 0.

This proves that c_p increases in l_r in the case of interior solutions. If the level of altruism a_p is so low that the contribution before and after the update is equal to zero ($\bar{c_p} = \ddot{c_p} = 0$), or if the contribution before is already at a maximum $\bar{c_p} = c^{max}$, the effect can be zero. Hence, the overall effect of an increase in l_r is non-negative on c_p. □

B.2.2 Proof of Proposition 2

Proof. To prove that the proposer's adjusted utility (v_p) increases in l_r, we show that a proposer can always choose a contribution $\ddot{c_p}$ that guarantees him a higher adjusted utility (v_p) than with a lower l_r. Following the experimental framework, we model an increase in l_r through learning the preferences of the matched receiver. This means that we demonstrate that a proposer's adjusted utility increases when he learns that l_r is higher than he previously thought. Note that we do not derive a proposer's optimal strategy, but show that there is always a strategy that makes the proposer better off.

The initially optimal contributions (given $\bar{l_r}$ and $\bar{a_p}$) by a proposer (receiver) are denoted by $\bar{c_p}$ ($\bar{c_r}$). The resulting monetary outcome of a proposer (receiver) is $\bar{m}_p$ ($\bar{m}_r$) and their direct (monetary) utility is $\bar{u_p}$ ($\bar{u_r}$). The preferences that the proposer then learns are denoted as $\ddot{l_r}$ ($> \bar{l_r}$), and the receiver's contribution is $\ddot{c_r}$ ($> \bar{c_r}$). The latter directly follows from Assumption 2. Note that if a player contributes c to the PGG, the sum of marginal returns for both players is greater than c. Therefore, contributing is always socially optimal.

In order to guarantee a higher adjusted utility, the proposer follows the following strategy: Contribute $\ddot{c_p}$, so that the receiver's new monetary outcome $\ddot{m_r}$ equals her old monetary outcome $\bar{m}_r$ (see Case 1). If this is not possible because it would require a higher contribution than is possible in the PGG ($\ddot{c}_p > c^{max}$), contribute the maximum possible contribution c^{max} to the PGG (see Case 2).

Case 1: Contribute $\ddot{c}_p$ such that $\ddot{m}_r = \bar{m}_r$.
The receiver's direct utility $\ddot{u}_r$ remains the same as her previous direct utility $\bar{u}_r$. Because both players contribute more, the overall monetary outcome is larger than before. Given that $\ddot{c}_p$ is set such that $\ddot{u}_r = \bar{u}_r$, the monetary payoff for proposer ($\ddot{m}_p$) must have increased. This implies that the proposer's adjusted utility must also increase, because his direct utility u_p and the level of altruism a_p increases, while the receiver's direct utility remains constant u_r.

This strategy might not always be possible. It can be the case that, even if the proposer contributes c^{max}, the new receiver's monetary outcome remains smaller than before ($\ddot{m}_r < \bar{m}_r$). Nevertheless, contributing c^{max} will always yield a a higher adjusted utility for the proposer v_p than before.

Case 2: Contribute $\ddot{c}_p = c^{max}$.
If the proposer contributes c^{max} and $\ddot{m}_r < \bar{m}_r$, the overall monetary outcome increases due to the increased overall contributions ($\ddot{m}_p + \ddot{m}_r > \bar{m}_p + \bar{m}_r$). Since $\ddot{m}_r < \bar{m}_r$, the monetary gain for the proposer must be greater than the monetary loss for the receiver ($\ddot{m}_p - \bar{m}_p > \bar{m}_r - \ddot{m}_r$). It must also follow that $\ddot{m}_r \geq \ddot{m}_p$ because the proposer contributes c^{max}. However, if the receiver also contributes $\ddot{c}_r = c^{max}$, both monetary outcomes are the same ($\ddot{m}_p = \ddot{m}_r$). Due to the concavity of the direct utility function, the increase in proposer's direct utility must be greater than the direct utility loss for the receiver. The increase of altruism even dampens the decrease of the receiver's direct utility u_r on the proposer's adjusted utility v_p. □

B.3 Exploratory Analyses

B.3.1 Determinants of Proposers' Preference Adjustments

Table B.5: Determinants of Proposers' Preference Adjustments

	1[Preference Adjustment]		
	No-Info	Info	
	(1)	(2)	(3)
Preference for initial partner (1-4)	-.122***	-.117***	-.124***
	[-.172,-.073]	[-.180,-.055]	[-.188,-.060]
Initial partner's preference (1-4)		-.081***	-.085***
		[-.140,-.021]	[-.145,-.025]
Average preference of other receivers (1-4)		.088**	
		[.010,.166]	
Highest preference of other receivers (1-4)			.092***
			[.041,.143]
Round	-.023**	-.027*	-.028*
	[-.045,-.001]	[-.056,.003]	[-.058,.002]
Loss Aversion	-.004	-.024	-.025
	[-.047,.040]	[-.081,.033]	[-.082,.033]
Cognitive Ability (Raven's)	.001	-.020	-.022
	[-.025,.028]	[-.080,.041]	[-.084,.039]
Male	.036	-.047	-.046
	[-.055,.127]	[-.165,.071]	[-.164,.073]
Observations	290	285	285

Notes. OLS Regressions. Standard errors are clustered at the individual level and reported in parentheses. *Preference for initial partner (1-4)* takes the value of four if the initial matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. *Round* is a count variable, indicating the number of the current round (Round 1-5). *Loss aversion* and *Cognitive ability* are calculated as detailed in Footnote 16, *Male* is an indicator taking the value of 1 if a participant indicated to identify as male. *Initial partner's preferences (1-4)* takes the value of four if the participant was the most preferred choice of their initial partner (i.e. before being able to adjust their preferences), three if the participant was the second most preferred choice, and so on. *Average preference of other receivers (1-4)* calculates the average preference of the other receiver and takes the value of four if the participant was the most preferred choice of all three receivers, the participant was not matched to initially. *Highest preference of other receivers (1-4)* takes the value of four if the partner was the most preferred choice of at least one of the non-matched receivers, three if the participant was not the most preferred choice of any receiver, but the second most preferred choice of at least one, and so on.

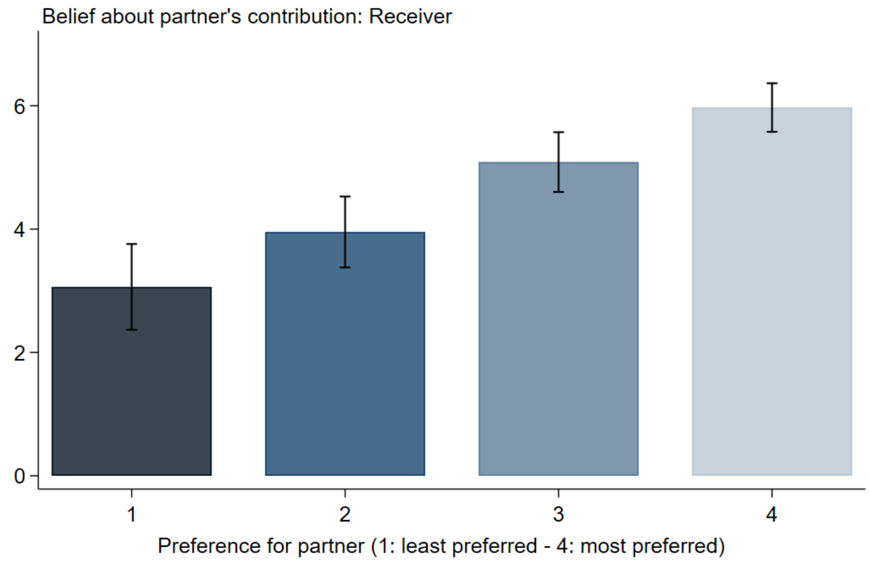
B.3.2 Predicting PGG Contribution with Questionnaire Responses

Table B.6: PGG Contributions and Questionnaire Responses

	Unconditional Contribution	Avg. Conditional Contribution
	(1)	(2)
Cat over Dog	-.207	.106
	[-.535,.121]	[-.204,.415]
Book over Film	.372	.005
	[-.119,.863]	[-.396,.406]
Beach over City	.150	-.028
	[-.343,.644]	[-.410,.353]
Bar over Club	-.176	-.225
	[-.678,.326]	[-.643,.192]
Living Alone over Shared	-.133	-.117
	[-.531,.264]	[-.487,.253]
Reserved	.455*	.179
	[-.025,.935]	[-.249,.607]
Lazy	.014	.021
	[-.509,.537]	[-.428,.470]
Handy with Hands	.261	.257
	[-.176,.698]	[-.105,.619]
Spontaneous	.092	.229
	[-.421,.605]	[-.290,.748]
Conflict Avoidant	.046	.227
	[-.456,.547]	[-.175,.629]
Strictness Covid19 Policy	-.108	.333
	[-.691,.475]	[-.097,.763]
Quota Disadvantaged	.417	-.031
	[-.081,.914]	[-.449,.387]
Bicycle Helmet Mandatory	.032	.055
	[-.420,.485]	[-.318,.428]
Legalize Marijuana	.342	.194
	[-.092,.775]	[-.219,.606]
Taxes Unhealthy Food	-.124	.106
	[-.543,.296]	[-.229,.441]
Observations	1150	575

Notes. OLS Regressions. Standard errors are clustered at the individual level and reported in parentheses. Column (1) includes both receivers and proposers. Column (2) only includes proposers, because receivers did not make conditional contribution decisions. For the wording of the questions, answered on a Likert scale from 1-4, see Appendix B.5.2.

B.3.3 Beliefs of Receivers about PGG Contribution of Partner



Notes. This figure displays the beliefs of receivers about the unconditional PGG contributions of their matched proposer by their preferences for the matched proposer. *Preference for partner (1-4)* takes the value of four if the matched proposer was the first choice of the receiver, three if the matched receiver was the second choice, and so on.

Figure B.2: Beliefs of Receivers: PGG Contributions of Partner

Table B.7: Beliefs of Receivers: PGG Contributions of Partner

	Beliefs about partner's PGG contribution (0-10)	
	(1)	(2)
Preference for partner (1-4)	.983***	.944***
	[.742,1.225]	[.708,1.181]
Round		-.057
		[-.198,.084]
Loss Aversion		-.234
		[-.710,.241]
Cognitive Ability (Raven's')		.213
		[-.111,.537]
Male		-.056
		[-1.142,1.029]
Observations	575	575

Notes. OLS Regressions. Standard errors are clustered at the individual level and reported in parentheses. *Preference for partner (1-4)* takes the value of four if the matched partner was the first choice of the participant, three if the matched partner was the second choice, and so on. *Round* is a count variable, indicating the number of the current round (Round 1-5). *Loss aversion* and *Cognitive ability* are calculated as detailed in Footnote 16, *Male* is an indicator taking the value of 1 if a participant indicated to identify as male.

B.3.4 Payoffs from PGG across Treatments

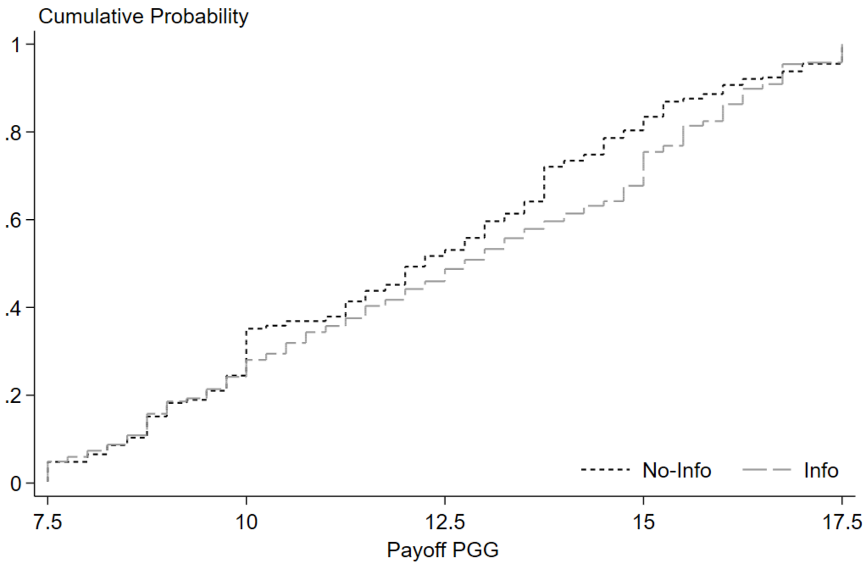


Figure B.3: Payoffs PGG: Implementation of Unconditional Decisions

B.4 Illustrative Example: Reciprocal Preferences and Stability in the DA

This example, borrowed from Opitz and Schwaiger (2023b), provides intuitions for the consequences of reciprocal preferences on stability in matching markets. It shows that the simultaneous DA mechanism (Gale & Shapley, 1962) can lead to an unstable allocation once (at least) one agent adjusts their preference order consistent with having reciprocal preferences..

We consider a job market with three firms (A, B, C), three workers (I, II, III), and a DA mechanism to match them one-to-one. All firms (A, B, C), as well as workers II and III have standard preferences. Both workers II and III prefer to only work for either of the firms over being unmatched. Worker II wants to work only for firm A, and worker III wants to work only for firm C. Worker I has reciprocal preferences: she cares how she is ranked by a firm. She prefers working for firm A over firm B if firm A ranks her first. If firm A ranks her second, then she prefers firm B over firm A. This means that her preference list is given by $A_1 \succ B \succ A_2$. The indices denote the true rank assigned to her by the respective firm. The (reciprocal) preferences of workers are common knowledge. Although the type of a firm is private knowledge, every agent knows the distribution of firms' types.

Acquiring information about the true preferences of firms is challenging in practice. If the firms' preferences were perfectly observable, then the preference list of worker I would reduce to a standard case. In reality, potential employees typically only have limited information about the exact demands of a firm and face uncertainty about the characteristics of the competing applicants. Moreover, employers may not be interested in truthfully revealing their preferences so that they can give each applicant the impression that they are a preferred candidate. Hence, we allow for uncertainty about the firms' preferences.

We incorporate uncertainty about the firms' preferences as follows. A firm knows its own realized type, but the other agents do not. In this example, firm A has two possible types denoted by a superscript A^1, A^2. Firm A^1 considers only worker I and II as potential employees, and has preferences of $I \succ II$. When being of type A^2, it only considers workers III and I, and prefers $III \succ I$. The probability of firm A being of type A^1 is p. Firm B only wants to be matched with worker I and firm C only wants to be matched with worker III. We summarize the information on the matching market displayed below. In addition, we assume that worker I has a higher expected utility of being matched with firm B than taking the lottery of being matched with firm A without knowing the type of firm A ($I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$).

Illustrative Example: Matching Market

Proposer / Firm		Receiver / Worker
$A^1 : I \succ II$	with (p)	$I : A_1 \succ B \succ A_2$
$A^2 : III \succ I$	with $(1-p)$	$II : A$
$B : I$		$III : C$
$C : III$		

Given:

$$I : u(B) > p \cdot u(A_1) + (1-p) \cdot u(A_2)$$

Workers can infer the type of firm A after observing the final match, given their knowledge of the matching market and mechanism. For example, if firm A is matched with worker II, then agents can infer that firm A is of type A^1 because type A^2 does not consider worker II as a relevant candidate. To illustrate the main intuitions, we first derive the optimal strategy of worker I in the DA mechanism, and then show that the outcome is unstable. Except for worker I, all agents in the matching market have standard preferences and will state these truthfully. Assuming that the utility of being matched with firm B is higher for worker I than the lottery of being matched with types A^1 or A^2, she states B ($\succ A$).

Given that worker I states $B \succ A$, the type of firm A will be revealed through the final matching. If type A^1 is realized, firm A is matched with worker II and if type A^2 is realized firm A is unmatched. Through observing the match of worker II and firm A, worker I can infer that firm A is of type A^1. Therefore, worker I wants to be matched with firm A and the matching is unstable. This happens because information about the type of a firm is revealed by the mechanism and the resulting final matching.

The same intuition holds for the experimental setting. The main difference is that the experiment does not require participants to make inferences about others' preferences through their knowledge about the mechanism and the matching outcome. In the experiment, we reveal the preferences of the other market side after the (tentative) matching takes place and give participants the option to adjust their preferences.

B.5 Instructions

Appendix B.5 includes the translated instructions of the experiment (from German). Treatment specific parts are shown in *italics* and the corresponding treatment is clearly indicated.

B.5.1 General Instructions (before Part I)

Welcome to the experiment and thank you for your participation!

Please do not speak from now on with any other participant.

Procedures

In this experiment, we study economic decision-making. You can earn money by participating. The money you earn will be paid to you privately after the experiment.
The experiment lasts around 90 minutes and consists of four parts (I-IV). At the beginning of every part, you receive detailed instructions. In addition, you will receive comprehension questions for some parts to help you understand how the experiment works and the payoff conditions. If you have questions after reading the instructions or during the experiment, please raise your hand or press the red button on your keyboard. One of the experimenters will then come to you and answer your questions privately.

Tools

You find a pen at your desk. Please leave the pen and the instructions on the table after the experiment.

Anonymity

The analysis of the experiment is anonymous; that is, we will never link your name with the data generated in the experiment. To receive your payoff, you will need to provide your bank details or PayPal mail address at the end of the experiment. No further personal data will be passed on. Information collected during the experiment may be visible to other participants as the experiment progresses. You make all decisions anonymously, so no other participant can associate your decisions with you during the experiment.

Payment
In addition to the income that you earn during the experiment, you will receive 6 € for showing up on time and answering a short questionnaire. In addition, you can achieve additional payoffs during the experiment. During the experiment, you and the other participants will be asked to make a series of decisions. These can affect the payoffs for you, and potentially for other participants. Additionally, you can earn money by making correct assessments. How your decisions relate to the payoffs will be explained in more detail in the respective instructions.

Exchange rate
In some parts of the experiment, we do not talk about Euro, but about Taler. We convert Taler into Euros at the end of the experiment. Please note the following exchange rate:

1 Taler = 0,70 €

B.5.2 Questionnaire (Part I)

[Instructions: *In the first part of the experiment, we ask you to truthfully fill out a questionnaire. This is a personality questionnaire, so there are no right or wrong answers. Please answer the questions with the answer options:*
• *Does not apply* • *Tends not to apply* • *Tends to apply* • *Applies*]

1. I would rather have a cat than a dog as a pet.
2. I prefer reading a book in the evening to watching a movie.
3. I prefer to go to the beach on vacation than to visit a city.
4. I would rather spend an evening in a bar than partying in a club.
5. I prefer to live in a shared apartment than alone.
6. I am rather reserved and quiet.
7. I am easygoing, prone to laziness.
8. I am talented with my hands.
9. I often make decisions spontaneously and intuitively.

10. I tend to avoid conflict.
11. I am in favor of strong policy measures to contain the Covid-19 pandemic in Germany.
12. I support quota regulations in the labor market for socially disadvantaged groups (e.g., for women or migrants).
13. There should be a requirement to wear a bicycle helmet.
14. The possession of marijuana should be legalized.
15. Unhealthy foods should be taxed more.

B.5.3 Instructions (Part II)

The participants received the instructions for Part II of the experiment in print. An interactive screen to familiarize with the matching procedure and control questions to ensure understanding were later displayed on the computer screens.

Proposer

Part II of the experiment consists of 5 rounds. Each round is structured in the same way. In each round, you will make decisions that affect your payout amount, as well as the payout amount of another participant. One round will be randomly selected for which the achieved amount will be paid out. You will find out which round was selected only at the end of the experiment. Therefore, you should carefully consider your decisions in all rounds, as each may become relevant to you.

You were randomly assigned one of two roles for Part II of the experiment. This role remains the same across all rounds. There are participants of "Type P" and participants of "Type R". You are "Type P". All participants of "Type P" receive identical instructions. Participants of "Type R" are in a similar decision situation, we explicitly point out any differences. In each round, four "Type P" participants are matched with four "Type R" participants. This means that 8 randomly selected participants interact with each other per round. In each round, you will be randomly selected to interact with other participants.

We will illustrate the process of Part II using one round as an example. We will refer to your group of four "Type P" participants as Group A, and to the group of four "Type R" participants with whom you interact as Group B.

Each round consists of three consecutive sections (Section 1, Section 2 and Section 3).

In the final Section 3, you will simultaneously make decisions with one participant from Group B (your team partner) that are payoff-relevant for both of you. In Section 3, one participant from Group A and one participant from Group B thus form a team of 2.

In Section 1, you specify which participant of Group B you want as your team partner in this decision situation. Your choice of team partner is important to you because your team partner's decisions affect your payoffs.

In Section 2, you will be assigned a team partner for Section 3 based on your choice and the choices of the other participants through an assignment mechanism.

Below you find detailed information on all three sections.

Section 1
In the first section, you will see a randomly selected part of the answers of the 4 participants of Group B from the questionnaire. These participants are your possible team partners.

Example image: Answers from the questionnaire

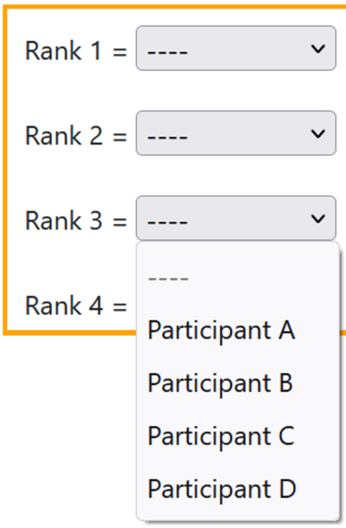
	Participant A	**Participant B**	**Participant C**	**Participant D**
Statement 1: I support quota regulations in the labor market for socially disadvantaged groups (e.g., for women or migrants).	Applies	Tends to apply	Tends not to apply	Tends not to apply
Statement 2: There should be a requirement to wear a bicycle helmet.	Does not apply	Applies	Does not apply	Does not apply
Statement 3: I would rather spend an evening in a bar than partying in a club.	Applies	Tends to apply	Does not apply	Applies
Statement 4: I am easygoing, prone to laziness.	Tends to apply	Tends to apply	Tends to apply	Does not apply
Aussage 5: I prefer to live in a shared apartment than alone.	Does not apply	Does not apply	Tends to apply	Tends to apply

At the same time, the participants of Group B (Participants A-D) see other randomly selected answers from your questionnaire and the questionnaires of the other 3 participants of Group A.

After viewing the profiles, we ask you to submit a preference order.

With this preference order, you indicate with whom of the participants from Group B you would prefer to be in the decision situation in Section 3. Rank 1 means that you would most like to have this participant as your team partner. Rank 2 means that you would second most like to have this participant as your team partner, and so on.

Example image: Preference order



All other participants of Groups A and B will also be asked to submit such a preference order.

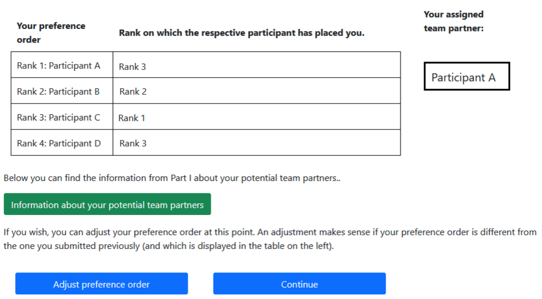
Section 2

In this section, a two-step mechanism will determine the allocation for Section 3. The mechanism is chosen so that it is always best for you to submit your actual preference order.

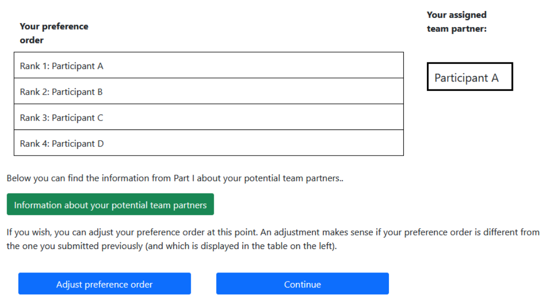
Example: Suppose you could choose between participants A, B, C or D from Group B. If you would prefer to have Participant A, second favorite Participant B, third favorite Participant C, and fourth favorite Participant D as your team partner, then you should submit the preference order A>B>C>D. If the assignment mechanism assigned you Participant B, for example, under the submission of your true preference order, there is no other preference order by which the mechanism assigns Participant A to you.
In the first step, the allocation mechanism determines the 2-person teams based on the preferences submitted. Then you will see which participant of Group B has been

assigned to you. *In addition, for each participant of Group B, you will see the rank they have placed you on. [Only in Info]*

Example screen: Adjustment of preferences



Info



No-Info

In this example, in the first step, you have set Participant A to Rank 1, and have been assigned him or her as a team partner by the mechanism. *Participant A has placed you on Rank 3 of their preference order. [Only in Info]*

If you wish, you can adjust your preference order at this point. An adjustment makes sense if your preference order is different from the one you submitted previously.

In the second step, the allocation mechanism again determines 2-person teams based on these preference orders. If at least one participant has adjusted their preference order, other teams may result compared to the teams after the first step. The key is that it is always best for you to submit your true preference order.

At the end of Section 2, it will be randomly selected whether your final team partner for Section 3 will be the one assigned to you after the first step, or whether your team partner will be the one assigned to you after the second step of Section 2. Therefore, you should submit your true preference order in both steps.

Information and procedure for participants of Group B

The process of Section 2 is different for participants from Group B. Unlike you, your potential team partners from Group B cannot adjust their preference order in the second part of the assignment mechanism. *Participants from Group B do not know the preference orders of Group A and do not know that Group A will receive the preference order of Group B. [Only in Info]*

Section 3

Decision situation

You and your team partner can each put 10 Taler into a private account, or you can put all or part of 10 Taler into a joint account. Any money that you do not deposit into the joint account will automatically be deposited into the private account. You and your team partner will make your decisions independently and secretly in this part.

Income from the private account

Every Taler you put on the private account, you will get paid at the end. If you keep 10 Taler for yourself, you will receive these 10 Taler from the private account. If you keep 6 Taler for yourself, you will receive these 6 Taler from the private account. Nobody but you receives income from your private account.

Income from the joint account

You can also put your Taler into the joint account. For each Taler contributed to the joint account, both you and your team partner will receive 0.75 Taler each. Both of you benefit from the joint account to the same extent, regardless of your respective deposits. The payoff from the joint account depends only on the sum of the deposits.

The payout of each team member is determined by the following formula.

Individual payout for each team member = (deposit from you + deposit from your team partner) * 0.75

If you and your team partner deposit 5 Taler each, the sum of the two deposits is 5+5=10. Of these 10 Taler, you and your team partner will each receive 10*0.75 = 7.5 Taler. If you and your team partner deposit a total of 16 Taler, you will both receive 16*0.75 = 12 Taler.

Total income

Your total income is the sum of your income from the personal account and your income from the joint account.

Your input

You and your team partner from Group B simultaneously and independently make the decision how many of your 10 Taler you want to contribute to the joint account. We call this decision contribution in the following.

In addition to this, participants in Group A make a second contribution decision, the contribution table. For participants of Group A, it is chosen at random whether the contribution or the contribution table is relevant for payout. You must therefore carefully consider both types of contribution decisions, as both may become relevant to you. Since participants of Group B only make the contribution decision, the contribution is always and exclusively payoff relevant for these participants.

Contribution and contribution table

With your contribution to the joint account, you determine how many of the 10 Taler you want to deposit into the joint account. The deposit to your private account is automatically the difference between 10 Taler and your contribution to the joint account.

Example image: Contribution

Please indicate the amount you wish to deposit into the joint account:

	0	1	2	3	4	5	6	7	8	9	10
How many of your 10 Taler do you contribute to the joint account?	○	○	○	○	○	○	○	○	○	○	○

In the contribution table, you specify how many Taler you want to contribute to the joint account for each possible contribution of your team partner. So you make your own contribution decision based on how much your team partner contributes.

Example image: Contribution table

For each possible contribution of your team partner, please indicate the amount you would like to contribute to the joint account (of course, you can choose the same amount more than once):

How many of your 10 Taler do you contribute to the joint account if....

	0	1	2	3	4	5	6	7	8	9	10
...Your team partner contributes 0 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 1 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 2 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 3 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 4 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 5 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 6 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 7 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 8 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 9 Taler?	○	○	○	○	○	○	○	○	○	○	○
...Your team partner contributes 10 Taler?	○	○	○	○	○	○	○	○	○	○	○

After the decision:

You will find out the result of the selected round only at the end of the experiment.

You can now familiarize yourself on the computer monitor with both the submission of preference sequences, as well as the allocation mechanism. After that, you will get some comprehension questions.

Receiver

Part II of the experiment consists of 5 rounds. Each round is structured in the same way. In each round, you will make decisions that affect your payout amount, as well as the payout amount of another participant. One round will be randomly selected for which the achieved amount will be paid out. You will find out which round was selected only at the end of the experiment. Therefore, you should carefully consider your decisions in all rounds, as each may become relevant to you.

You were randomly assigned one of two roles for Part II of the experiment. This role remains the same across all rounds. There are participants of "Type P" and participants of "Type R". You are "Type R". All participants of "Type R" receive identical instructions. Participants of "Type P" are in a similar decision situation. In each round, four "Type P" participants are matched with four "Type R" participants. This means that 8 randomly selected participants interact with each other per round. In each round, you will be randomly selected to interact with other participants.

We will illustrate the process of Part II using one round as an example. We will refer to your group of four "Type P" participants with whom you interact as Group A, and your group of four "Type R" participants as Group B.

Each round consists of three consecutive sections (Section 1, Section 2 and Section 3).

In the final Section 3, you will simultaneously make decisions with one participant from Group A (your team partner) that are payoff-relevant for both of you. In Section 3, one participant from Group A and one participant from Group B thus form a team of 2.

In Section 1, you specify which participant of Group A you want as your team partner in this decision situation. Your choice of team partner is important to you because your team partner's decisions affect your payoffs.

In Section 2, you will be assigned a team partner for Section 3 based on your choice and the choices of the other participants through an assignment mechanism.

Below you find detailed information on all three sections.

Section 1
In the first section, you will see a randomly selected part of the answers of the 4 participants of Group A from the questionnaire. These participants are your possible team partners.

Example image: Answers from the questionnaire

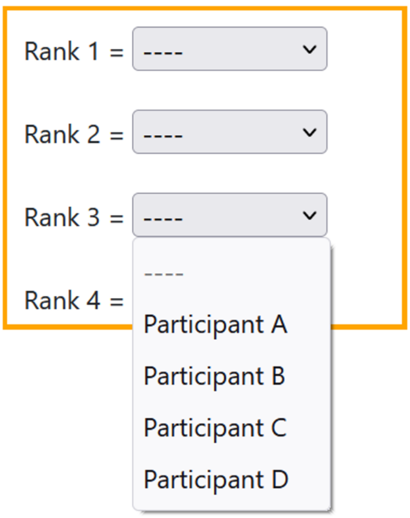
	Participant A	Participant B	Participant C	Participant D
Statement 1: I support quota regulations in the labor market for socially disadvantaged groups (e.g., for women or migrants).	Applies	Tends to apply	Tends not to apply	Tends not to apply
Statement 2: There should be a requirement to wear a bicycle helmet.	Does not apply	Applies	Does not apply	Does not apply
Statement 3: I would rather spend an evening in a bar than partying in a club.	Applies	Tends to apply	Does not apply	Applies
Statement 4: I am easygoing, prone to laziness.	Tends to apply	Tends to apply	Tends to apply	Does not apply
Aussage 5: I prefer to live in a shared apartment than alone.	Does not apply	Does not apply	Tends to apply	Tends to apply

At the same time, the participants of Group A (Participants A-D) see other randomly selected answers from your questionnaire and the questionnaires of the other 3 participants of Group B.

After viewing the profiles, we ask you to submit a preference order.

With this preference order, you indicate with whom of the participants from Group A you would prefer to be in the decision situation in Section 3. Rank 1 means that you would most like to have this participant as your team partner. Rank 2 means that you would second most like to have this participant as your team partner, and so on.

Example image: Preference order



All other participants of Groups A and B will also be asked to submit such a preference order.

Section 2

In this section, a mechanism will determine the allocation for Section 3. The goal of the mechanism is to assign participants their best possible team partner. The mechanism is based on a simple logic: If several participants of Group A want you to be their team partner, the mechanism will always select for you the participant that you have specified further ahead in your preference order.

Example: Suppose you could choose between participants A, B, C or D from Group A. You prefer to have Participant A, second favorite Participant B, third favorite Participant C, and fourth favorite Participant D as your team partner (A>B>C>D). If the assignment mechanism does not assign you Participant A when you state your true preference order, it automatically means that Participant A prefers another participant of Group B over you.

Let us assume that this is the case. Now, if both participant B and C would prefer you to be their team partner, the mechanism will choose the participant you have specified further up in your preference order as your team partner. If you would submit the preference order A>B>C>D, you would get Participant B as your team partner. If you would give the preference order A>C>B>D, you would get Participant C as your team partner. This also means that if you submit a preference order that does not match your true preference order, you may not get your best possible team partner.

Once you have submitted your preference order, you cannot change it.

Section 3
Decision situation

You and your team partner can each put 10 Taler into a private account, or you can put all or part of 10 Taler into a joint account. Any money that you do not deposit into the joint account will automatically be deposited into the private account. You and your team partner will make your decisions independently and secretly in this part.

Income from the private account

Every Taler you put on the private account, you will get paid at the end. If you keep 10 Taler for yourself, you will receive these 10 Taler from the private account. If you keep 6 Taler for yourself, you will receive these 6 Taler from the private account. Nobody but you receives income from your private account.

Income from the joint account

You can also put your Taler into the joint account. For each Taler contributed to the joint account, both you and your team partner will receive 0.75 Taler each. Both of you benefit from the joint account to the same extent, regardless of your respective deposits. The payoff from the joint account depends only on the sum of the deposits.

The payout of each team member is determined by the following formula.

Individual payout for each team member =
(deposit from you + deposit from your team partner) * 0.75

If you and your team partner deposit 5 Taler each, the sum of the two deposits is 5+5=10. Of these 10 Taler, you and your team partner will each receive 10*0.75 = 7.5 Taler. If you and your team partner deposit a total of 16 Taler, you will both receive 16*0.75 = 12 Taler.

Total income

Your total income is the sum of your income from the personal account and your income from the joint account.

Your input

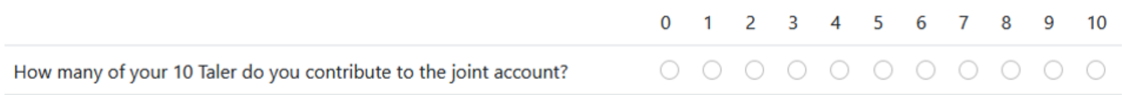
You and your team partner from Group B simultaneously and independently make the decision how many of your 10 Taler you want to contribute to the joint account. We call this decision contribution in the following.

Contribution

With your contribution to the joint account, you determine how many of the 10 Taler you want to deposit into the joint account. The deposit to your private account is automatically the difference between 10 Taler and your contribution to the joint account.

Example image: Contribution

Please indicate the amount you wish to deposit into the joint account:



After the decision:

You will find out the result of the selected round only at the end of the experiment.

You can now familiarize yourself on the computer monitor with both the submission of preference sequences, as well as the allocation mechanism. After that, you will get some comprehension questions.

B.5.4 Additional Instructions (Part III)

Beliefs

In this part of the experiment, we ask you to guess the decisions of your respective team partners from Part II. You thus provide an estimate for each of the rounds played. Your payoff depends on whether you estimate the contribution to the joint account of your respective team partner in Part II correctly.
Before each decision, you will again receive the information about your team partner that you had available when you made your own contribution decision. Please provide an estimate of how many Taler your respective team partner put into the joint account. *Note that your team partner made this decision, without knowing your submitted preference order. [only proposer]*
Payoff
If you estimate your team partner's contribution exactly correctly, you will receive 2 Euro for this correct estimation. If you estimate the contribution incorrectly, you will receive 0 Euro.
One of the rounds will be randomly selected for which the amount scored will be paid out. You will find out the result of the selected round only at the end of the experiment (after part IV).

Raven's Matrices

In this part of the experiment we ask you to complete figures. The figures consist of 3x3 elements that are logically connected. In each figure the lower right element is missing. We ask you to complete this with one of the 6 answer choices.
You have a total of 5 minutes to solve as many matrices as you can manage. The maximum number is 10 matrices. You will receive 0.50 Euro for each correctly solved matrix and 0.50 Euro will be deducted for each incorrectly solved matrix. You will receive at least 0.00 Euro for this task. You cannot get a negative payout from this task. Please select the appropriate image in each case and confirm your selection. On the next page you can see an example.

Loss attitudes (Gächter, Johnson, & Herrmann, 2022)

This task consists of 6 decisions where you can accept up to 6 offers.
The offers consist of a lottery through which you can lose or win money. You have to decide for each of the 6 offers whether to accept it or not. For each accepted offer,

the computer plays the lottery and hence decides if you lose or win money.
At the end of the experiment, your decision is implemented for one of the 6 offers. The computer randomly selects (with equal probability) which offer will be implemented.

Decide for each offer whether you want to accept it.

1	With 50% probability you lose 2 Euro; with 50% probability you win 6 Euro.	∘ accept ∘ reject
2	With 50% probability you lose 3 Euro; with 50% probability you win 6 Euro.	∘ accept ∘ reject
3	With 50% probability you lose 4 Euro; with 50% probability you win 6 Euro.	∘ accept ∘ reject
4	With 50% probability you lose 5 Euro; with 50% probability you win 6 Euro.	∘ accept ∘ reject
5	With 50% probability you lose 6 Euro; with 50% probability you win 6 Euro.	∘ accept ∘ reject
6	With 50% probability you lose 7 Euro; with 50% probability you win 6 Euro.	∘ accept ∘ reject

Socio-demographics

Please provide the following statistical information.

- Age [integer]
- Gender [male; female; diverse]
- Field of study (faculty/major) [string]
- What language(s) is (are) your native language(s)? [string]
- What is your high school graduation grade? [number; 1-6]
- What is your high school graduation grade in mathematics? [number; 1-6]
- How many times have you participated in an economic laboratory study (including outside of this laboratory)? [0; 1-2; 3-5; 5+]

C

Appendix to Chapter 3

Interpersonal Preferences and Team Performance

C.1 Design, Questionnaires & Materials

C.1.1 Timeline of the Experiment

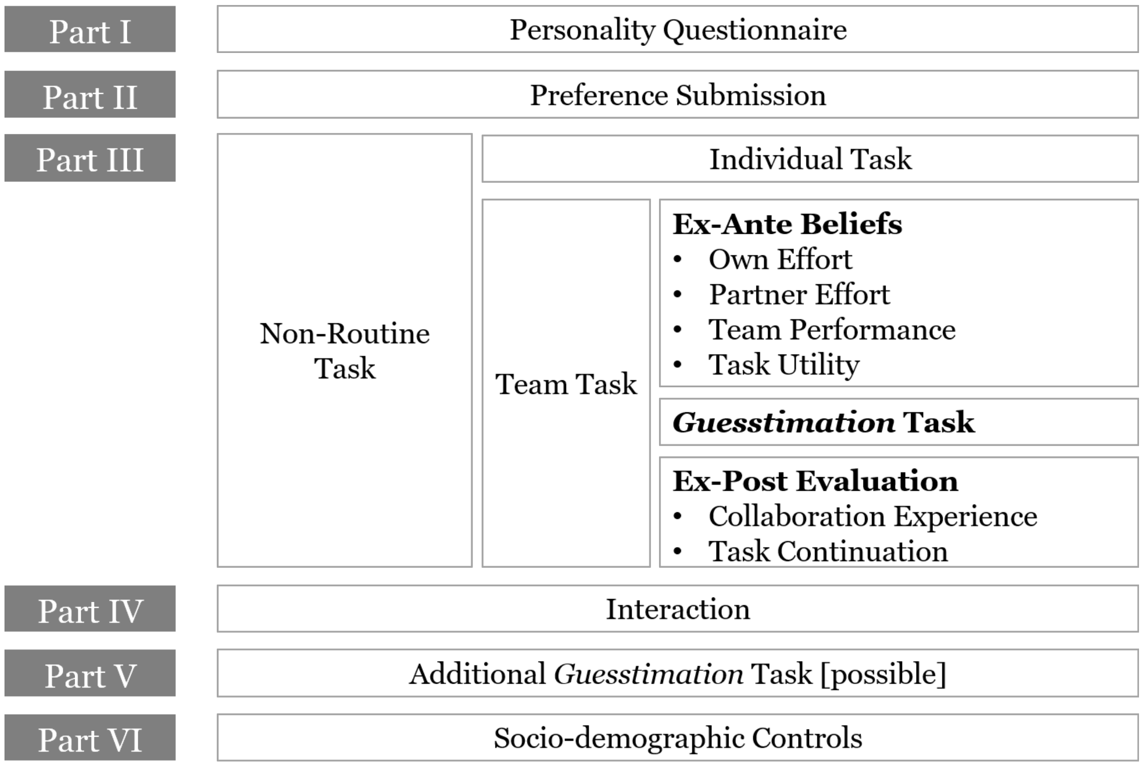


Figure C.1: Timeline of the Experiment

C.1.2 Personality Questionnaire

1. I would rather have a cat than a dog as a pet.
2. I prefer reading a book in the evening to watching a movie.
3. I prefer going to the beach on vacation to visiting a city.
4. I would rather spend an evening in a bar than partying in a club.
5. I prefer to living in a shared apartment than alone.
6. I am rather reserved and quiet.
7. I am easygoing, prone to laziness.

8. I am talented with my hands.
9. I often make decisions spontaneously and intuitively.
10. I tend to avoid conflict.
11. I support climate protests, even if they use means of civil disobedience (e.g., “Last Generation”).
12. I support quota regulations in the labor market for socially disadvantaged groups (e.g., for women or migrants).
13. There should be a requirement to wear a bicycle helmet.
14. The possession of marijuana should be legalized.
15. Unhealthy foods should be taxed more.

C.1.3 *Guesstimation* Tasks

In the experiment, each team faces one of the following *Guesstimation* tasks.

1. What is the route length of all streetcar lines in Germany?
2. How many businesses in the hairdressing trade (hair salons) were there in Germany in 2022?
3. How many dogs are there in Germany?
4. How much household waste was generated per inhabitant in Germany in 2021 (household waste + bulky waste + recyclables)?
5. How many passengers were transported by Deutsche Bahn ICEs in 2022?

C.1.4 Collaboration Questionnaire

Psychological Safety (Edmondson, 1999, 2002)

- I felt like my team partner would judge me on the things that I said.
- I feel safe sharing my views with my team partner.

Task Satisfaction (Dimotakis, Davison, & Hollenbeck, 2012)

- I found real enjoyment in performing this task.

Group Climate (Van Ginkel & Van Knippenberg, 2008)

- The atmosphere was good
- Our collaboration was good

Willingness to Compromise (De Dreu et al., 2001)

- I tried to accommodate my teammate.
- I insisted we both give in a little.

Fear of Disappointing Others

- I did not want to disappoint my team partner.

C.2 Additional Analyses

C.2.1 Treatment *No-Info*: Participants do not know each others' preferences

There are no significant effects of team partners' preferences on team performance in *No-Info*. This confirm that not team characteristics are underlying the effects in the treatment *Info*. Also within *No-Info*, I compare teams with different preference constellations –with the only differences that participants were not aware of these preferences. Columns (1)-(4) of Table C.1 estimate the pre-registered specifications, Column (5) adds *Guesstimation* problem fixed effects. Results consistently show that there is no differential performance depending on the preferences of both partners. The positive point estimates for *joint liking* at least shows that there is no negative performance of sorting. While participants generally prefer to interact with those who are similar to them ($r = 0.119$, $p < 0.001$; Pearson's correlation coefficient based on the similarity of responses to five the questions that participants saw and *liking*), this does not translate into worse performance due to a lack in complementary skills or knowledge.

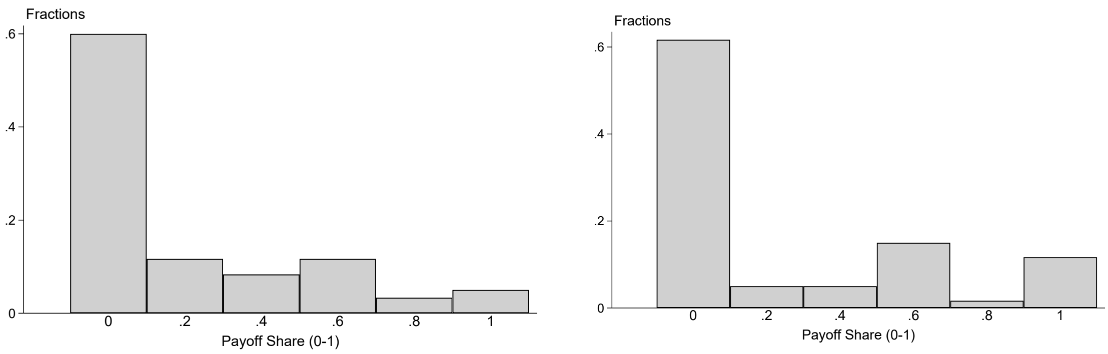
Table C.1: Task Performance: Guesstimation [*No-Info*]

	Payoff Share (0-1)				
	(1)	(2)	(3)	(4)	(5)
Joint Liking (2-8)	.038	.031	.039	.031	.032
	[-.011,.088]	[-.022,.083]	[-.011,.089]	[-.022,.084]	[-.023,.088]
Dissimilarity Liking (0-3)			-.029	-.028	-.017
			[-.142,.083]	[-.151,.095]	[-.148,.115]
Mean Dep. Var.	.25	.25	.25	.25	.25
Team Controls	No	Yes	No	Yes	Yes
Performance Controls	No	Yes	No	Yes	Yes
Problem FE	No	No	No	No	Yes
N	60	60	60	60	60

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is the fraction of the maximum possible payoff that teams achieved. The values in square brackets represent the 95% confidence intervals. *Joint Liking* is the sum of both partners' preferences. *Dissimilarity Liking* is the difference between both partners' preferences. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants could face.

C.2.2 Task Performance across Treatments

Task performance is similar across *Info* and *No-Info*. The average payoff share is 0.25 in *Info* and 0.20 in *No-Info* ($p = 0.765$; Mann-Whitney-U test (MWU)). Also both distributions (see Figure C.2) are not statistically significantly different from each other ($p = 0.985$; Kolmogorov–Smirnov test). Therefore, on average, there are no efficiency gains from knowing the preferences of one's partner before performing the *Guesstimation* task.



A) Payoff Shares in *Info*

B) Payoff Shares in *No-Info*

Notes. This figure displays distributions of payoff share in the joint *Guesstimation* task. Panel A shows the distributions in *Info*, Panel B in *No-Info*.

Figure C.2: Payoff Share across Treatments

C.2.3 Ex-ante Beliefs and Preferences

Beliefs: Own Effort

Table C.2: Own Effort [*Info*]

	Expected Own Effort (0-10)			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	.131			
	[-.094,.356]			
Partner's Liking (1-4)		.060		
		[-.171,.291]		
Joint Liking (2-8)			.087	.091
			[-.066,.240]	[-.063,.245]
Dissimilarity Liking (0-3)				-.080
				[-.357,.197]
Mean Dep. Var.	8.692	8.692	8.692	8.692
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
N	120	120	120	120

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is self-reported motivation to exert high effort. The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Table C.3: Own Effort [across Treatments]

	Expected Own Effort (0-10)			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	.093			
	[-.154,.339]			
Own Liking X Info	.042			
	[-.299,.382]			
Partner's Liking (1-4)		.077		
		[-.171,.324]		
Partner's Liking X Info		-.026		
		[-.368,.315]		
Joint Liking (2-8)			.056	.054
			[-.086,.199]	[-.090,.197]
Joint Liking X Info			.027	.033
			[-.186,.240]	[-.180,.247]
Dissimilarity Liking (0-3)				.109
				[-.209,.427]
Dissimilarity Liking X Info				-.188
				[-.614,.238]
Info	-.045	.120	-.074	.073
	[-.942,.853]	[-.780,1.021]	[-1.161,1.012]	[-1.069,1.216]
Mean Dep. Var.	8.65	8.65	8.65	8.65
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
Problem FE	No	No	No	No
N	240	240	240	240

Notes. OLS Regressions.*** $p<0.01$, **$p<0.05$, * $p<0.1$. The outcome variable is self-reported motivation to exert high effort. The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Own Liking X Info* interacts this variable with a treatment indicator for *Info*; the other interactions in the table are defined accordingly. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on.*Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Beliefs: Partner Effort

Table C.4: Partner Effort [*Info*]

	Expected Partner Effort			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	.199			
	[-.050,.448]			
Partner's Liking (1-4)		.299**		
		[.048,.550]		
Joint Liking (2-8)			.224***	.229***
			[.058,.390]	[.062,.396]
Dissimilarity Liking (0-3)				-.101
				[-.401,.200]
Mean Dep. Var.	8.40	8.40	8.40	8.40
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
N	120	120	120	120

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is the belief about the partner's motivation to exert high effort. The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Own Liking X Info* interacts this variable with a treatment indicator for *Info*; the other interactions in the table are defined accordingly. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on.*Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Table C.5: Partner Effort [across Treatments]

	Expected Partner Effort (0-10)			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	.241*			
	[-.034,.516]			
Own Liking X Info	-.053			
	[-.432,.326]			
Partner's Liking (1-4)		.009		
		[-.266,.284]		
Partner's Liking X Info		.273		
		[-.107,.652]		
Joint Liking (2-8)			.083	.082
			[-.075,.241]	[-.077,.241]
Joint Liking X Info			.129	.137
			[-.107,.364]	[-.100,.374]
Dissimilarity Liking (0-3)				.057
				[-.295,.409]
Dissimilarity Liking X Info				-.186
				[-.658,.286]
Info	.034	-.742	-.712	-.549
	[-.965,1.033]	[-1.742,.258]	[-1.916,.492]	[-1.815,.718]
Mean Dep. Var.	8.43	8.43	8.43	8.43
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
Problem FE	No	No	No	No
N	240	240	240	240

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is the belief about the partner's motivation to exert high effort. The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Own Liking X Info* interacts this variable with a treatment indicator for *Info*; the other interactions in the table are defined accordingly. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on.*Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Beliefs: Team Performance

Table C.6: Team Success [across Treatments]

	Expected Team Performance			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	-.443*			
	[-.923,.037]			
Own Liking X Info	1.032***			
	[.371,1.694]			
Partner's Liking (1-4)		-.104		
		[-.591,.383]		
Partner's Liking X Info		.501		
		[-.170,1.172]		
Joint Liking (2-8)			-.185	-.176
			[-.462,.093]	[-.454,.102]
Joint Liking X Info			.629***	.615***
			[.216,1.042]	[.201,1.029]
Dissimilarity Liking (0-3)				-.400
				[-1.016,.215]
Dissimilarity Liking X Info				.503
				[-.323,1.328]
Info	-2.636***	-1.367	-3.186***	-3.543***
	[-4.381,-.892]	[-3.137,.402]	[-5.295,-1.076]	[-5.758,-1.329]
Mean Dep. Var.	7.43	7.43	7.43	7.43
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
Problem FE	No	No	No	No
N	240	240	240	240

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is the amount bet on the team being in the top 25%, with monetary values recoded on a scale of 0-10.The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Own Liking X Info* interacts this variable with a treatment indicator for *Info*; the other interactions in the table are defined accordingly. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on.*Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Beliefs: Task Utility

Table C.7: Task Utility [across Treatments]

	Preference for Working Alone			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	.032			
	[-.337,.401]			
Own Liking X Info	-.047			
	[-.556,.462]			
Partner's Liking (1-4)		.151		
		[-.218,.519]		
Partner's Liking X Info		-.128		
		[-.636,.380]		
Joint Liking (2-8)			.061	.063
			[-.152,.275]	[-.151,.277]
Joint Liking X Info			-.058	-.074
			[-.375,.260]	[-.393,.244]
Dissimilarity Liking (0-3)				-.055
				[-.529,.418]
Dissimilarity Liking X Info				.352
				[-.282,.987]
Info	.374	.565	.536	.208
	[-.968,1.716]	[-.775,1.905]	[-1.087,2.159]	[-1.494,1.910]
Mean Dep. Var.	5.34	5.34	5.34	5.34
Team Controls	Yes	Yes	Yes	
Performance Controls	Yes	Yes	Yes	Yes
Problem FE	No	No	No	No
N	240	240	240	240

Notes. OLS Regressions.*** p <0.01, ***p* <0.05, * p <0.1. The outcome variable indicates the willingness to pay to perform another round of the task with a fixed payoff alone (as opposed to with their team partner). The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Own Liking X Info* interacts this variable with a treatment indicator for *Info*; the other interactions in the table are defined accordingly. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on.*Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

C.2.4 Communication

Table C.8: Team Communication [*Info*]

	Total Length of Chat Messages				
	(1)	(2)	(3)	(4)	(5)
Joint Liking (2-8)	18.3	12.9	21.3	17.8	12.6
	[-32.8,69.5]	[-38.3,64.1]	[-29.5,72.1]	[-32.6,68.2]	[-41.5,66.6]
Dissimilarity Liking (0-3)			-65.8	-85.5*	-77.8
			[-156.1,24.6]	[-178.9,7.9]	[-185.1,29.6]
Mean Dep. Var.	906.6	906.6	906.6	906.6	906.6
Team Controls	No	Yes	No	Yes	Yes
Performance Controls	No	Yes	No	Yes	Yes
Problem FE	No	No	No	No	Yes
N	60	60	60	60	60

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The values in square brackets represent the 95% confidence intervals. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants face.

Table C.9: Team Communication [across Treatments]

	Asymmetric Communication	Turn-Taking	Time to First Message
	(1)	(2)	(3)
Joint Liking (2-8)	.001	-.432	-.240
	[-.012,.013]	[-1.473,.608]	[-3.793,3.313]
Joint Liking X Info	-.007	.502	-2.057
	[-.026,.011]	[-1.068,2.072]	[-7.418,3.303]
Dissimilarity Liking (0-3)	.003	-.401	2.544
	[-.025,.032]	[-2.789,1.988]	[-5.613,10.702]
Dissimilarity Liking X Info	.021	-.842	1.516
	[-.017,.058]	[-4.049,2.365]	[-9.437,12.468]
Info	-.010	-1.725	5.347
	[-.112,.091]	[-10.315,6.866]	[-23.988,34.681]
Mean Dep. Var.	.11	19.55	41.21
Team Controls	Yes	Yes	Yes
Performance Controls	Yes	Yes	Yes
Problem FE	Yes	Yes	Yes
N	120	120	120

Notes. OLS Regressions.*** p <T0.01, **p <0.05, * p <0.1. The values in square brackets represent the 95% confidence intervals. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Joint Liking X Info* interacts this variable with a treatment indicator for *Info*. *Dissimilarity Liking* is the absolute difference between both partners' preferences, *Dissimilarity Liking X Info* captures the interaction with the treatment *Info*. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants face.

C.2.5 Ex-post Evaluation

Collaboration Experience

Table C.10: Collaboration Experience [*Info*]

	Regression Coefficient [Confidence Interval]	
	Sum of partner's preferences	Preference Difference
Psychological Safety	-.019 [-.097,.059]	-.005 [-.145,.136]
Group Climate	.030 [-.079,.139]	-.036 [-.232,.161]
Task Enjoyment	.077 [-.058,.211]	.090 [-.151,.331]
Willingness to Compromise	.057 [-.032,.146]	-.101 [-.260,.059]
Fear of Disappointing Others	.020 [-.088,.128]	-.052 [-.246,.142]

Notes. OLS Regressions. *** p <0.01, ** p <0.05, * p <0.1. The values in square brackets represent the 95% confidence intervals. This table displays coefficients and confidence intervals from individual regressions, each controlling for age, gender, and individual performance.

Task Continuation

Table C.11: Task Continuation [*Info*]

	Preference for Working Alone			
	(1)	(2)	(3)	(4)
Own Liking (1-4)	.019			
	[-.059,.096]			
Partner's Liking (1-4)		-.010		
		[-.089,.069]		
Joint Liking (2-8)			.004	.004
			[-.049,.057]	[-.050,.057]
Dissimilarity Liking (0-3)				.007
				[-.088,.103]
Mean Dep. Var.	.40	.40	.40	.40
Demographic Controls	Yes	Yes	Yes	Yes
Performance Indiv. Task	Yes	Yes	Yes	Yes
N	120	120	120	120

Notes. OLS Regressions.*** p <0.01, **p <0.05, * p <0.1. The outcome variable is a (binary) indicator of whether the participant would like to perform another round of the task alone (as opposed to with their team partner). The values in square brackets represent the 95% confidence intervals. *Own Liking (1-4)* takes the value of four if the partner was the first choice of the participant, three if the partner was the second choice, and so on. *Partner's Liking (1-4)* takes the value of four if the participant was the most preferred choice of their partner, three if the participant was the second most preferred choice, and so on. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Dissimilarity Liking* is the absolute difference between both partners' preferences. The *Demographic Controls* are the age and gender of the participant, *Performance Indiv. Task*, captures their performance in the individual *Guesstimation* task.

Correlation Table

Table C.12: Correlations of Message Length and Collaboration Experience

	Communication Quantity (1)	Psychological Safety (2)	Fear of Disappointing (3)	Task Enjoyment (4)	Willingness to Compromise (5)	Group Climate (6)
Communication Quantity	1.00					
Psychological Safety	0.01	1.00				
Fear of Disappointing	−0.02	−0.16 *	1.00			
Task Enjoyment	0.06	0.36 ***	−0.02	1.00		
Willingness to Compromise	0.18 *	0.21 **	−0.03	0.05	1.00	
Group Climate	0.12	0.40 ***	−0.07	0.56 ***	0.09	1.00

Notes. Pearson's Correlation Coefficients. *** p <0.01, ** p <0.05, * p <0.1.

C.2.6 Adjusted Performance Measure

Table C.13: Task Performance: Guesstimation [across Treatments]

	Payoff Share (0-1)		
	(1)	(2)	(3)
Joint Liking (2-8)	.022	.023	.026
	[-.025,.069]	[-.023,.070]	[-.022,.073]
Joint Liking X Info	-.059*	-.066*	-.067*
	[-.129,.011]	[-.135,.003]	[-.139,.004]
Info	.253	.126	.137
	[-.105,.611]	[-.244,.496]	[-.254,.529]
Dissimilarity Liking (0-3)		-.054	-.049
		[-.160,.052]	[-.158,.060]
Dissimilarity Liking X Info		.150**	.151**
		[.010,.290]	[.005,.297]
Mean Dep. Var.	.25	.25	.25
Team Controls	Yes	Yes	Yes
Performance Controls	Yes	Yes	Yes
Problem FE	No	No	Yes
N	120	120	120

Notes. OLS Regressions.*** p <0.01, ** p <0.05, * p <0.1. The values in square brackets represent the 95% confidence intervals. The outcome variable is the fraction of the maximum possible payoff that teams would have achieved when correcting for typos or formatting errors team members made. *Joint Liking* is the sum of both partners' preferences. For example, a value of 8 means that both partners ranked the other as their first choice, a value of 2 means that both partners ranked the other as their last choice. *Joint Liking X Info* interacts this variable with a treatment indicator for *Info*. *Dissimilarity Liking* is the absolute difference between both partners' preferences, *Joint Liking X Info* captures the interaction with the treatment *Info*. The *Team Controls* are the age and gender of both team partners, *Performance Controls*, capture the individual task performance of each partner. *Problem FE* are fixed effects for the five different *Guesstimation* problems that participants face.

C.2.7 Predictive Potential Questionnaire

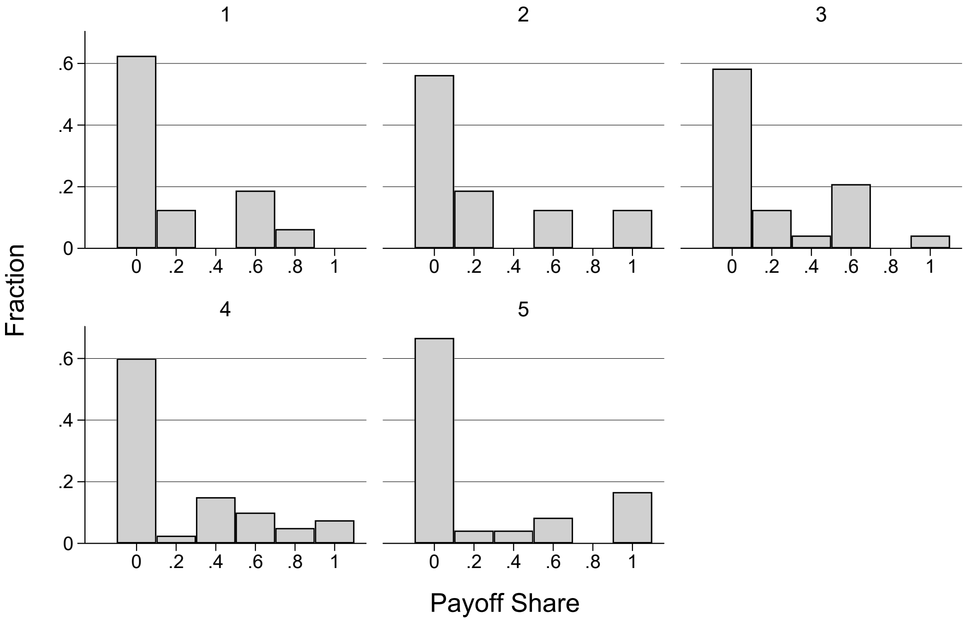
Predictive Potential Questionnaire

Table C.14: Predictive Potential Questionnaire

	Individual Performance	Joint Performance	
	All	All	Info
	(1)	(2)	(3)
Cat over Dog	-.000	-.003	.008
	[-.029,.029]	[-.039,.032]	[-.042,.059]
Book over Film	-.022	-.043	-.062
	[-.066,.022]	[-.097,.010]	[-.138,.013]
Beach over City	-.002	.026	.040
	[-.040,.036]	[-.020,.073]	[-.027,.107]
Bar over Club	-.010	.008	.016
	[-.049,.029]	[-.040,.055]	[-.051,.083]
Living Alone over Shared	.028	.024	.009
	[-.006,.061]	[-.018,.065]	[-.048,.067]
Reserved	.031	-.018	-.032
	[-.009,.071]	[-.067,.032]	[-.102,.037]
Lazy	-.031	-.038	-.034
	[-.074,.011]	[-.090,.014]	[-.112,.043]
Handy with Hands	.035*	-.037	.011
	[-.006,.076]	[-.088,.013]	[-.061,.084]
Spontaneous	-.005	.055**	.038
	[-.050,.039]	[.001,.110]	[-.039,.115]
Conflict Avoidant	.024	.038	-.022
	[-.018,.066]	[-.014,.090]	[-.090,.046]
Climate Civil Disobedience	.006	-.029	-.016
	[-.039,.051]	[-.084,.026]	[-.093,.060]
Quota Disadvantaged	-.020	.013	-.030
	[-.063,.022]	[-.039,.065]	[-.103,.042]
Bicycle Helmet Mandatory	-.000	.017	.066*
	[-.040,.040]	[-.031,.066]	[-.003,.135]
Legalize Marijuana	.007	.014	.019
	[-.029,.044]	[-.031,.058]	[-.044,.081]
Taxes Unhealthy Food	.008	-.014	.010
	[-.029,.046]	[-.060,.031]	[-.049,.069]
Observations	240	240	120

Notes. OLS Regressions. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The values in square brackets represent the 95% confidence intervals. Column (1) estimates the effect of questionnaire responses on the individual task performance. Columns (2) & (3) estimate the effect on the performance in the joint task. Column (2) includes all, participants, Column (3) only includes participants in the *Info* condition. For the wording of the questions, answered on a Likert scale from 1-4, see Appendix C.1.2.

C.2.8 Performance by *Guesstimation Task*



Notes. This figure displays the payoff shares across the five different *Guesstimation* tasks. Task 1 asked about the length of all streetcar lines in Germany, Task 2 about the number of hair salons in Germany, Task 3 about the number of dogs in Germany, Task 4 about the yearly household waste of an inhabitant in Germany, and Task 5 about the number of passengers transported by Deutsche Bahn ICEs. All questions refer to numbers in either 2021, 2022, or 2023.

Figure C.3: Payoff Share across *Guesstimation* Problems

C.3 Instructions

Appendix C.3 includes the translated instructions of the experiment (from German). Treatment specific parts are shown in *italics* and the corresponding treatment is clearly indicated.

General (On-screen) Instructions (before Part I)

Welcome to the experiment and thank you for your participation!

Please do not speak from now on with any other participant.

Procedures

This experiment is designed to study economic decision-making. You can earn money by participating. The money you earn will be paid to you privately and in cash after the experiment.

The experiment lasts around 60 minutes and consists of five parts (I-V). At the beginning of every part, you receive detailed instructions. In addition, you will receive comprehension questions for some parts to help you understand how the experiment and the payoff conditions work. If you have questions after reading the instructions or during the experiment, please raise your hand or press the red button on your keyboard. The experimenter will then come to you and answer your questions privately.

Anonymity

The analysis of the experiment is anonymous; that is, we will never link your name with the data generated in the experiment. At the end of the experiment, you must sign a receipt for the receipt of the earnings. This receipt is only used for accounting and booking of the experiment money. No further personal data from the experiment will be passed on. Information collected during the experiment may be visible to other participants as the experiment progresses. You make all decisions anonymously, so no other participant can associate your decisions with you during the experiment.

Payment

For your participation in this experiment and the completion of a short questionnaire at the end of it you receive 10 €. In addition, you can achieve further payoffs during the experiment. During the experiment, you and the other participants will be asked to make a series of decisions. These can affect the payoffs for you, and potentially for other participants. Additionally, you can earn money by making correct assessments. You will receive detailed instructions about these decisions and how they affect the payoff.

Information on fold-out elements
On some places you will find green buttons. When you click on them, they unfold and contain information. Here you can see an example:

Sample Box

Blue buttons do not contain any information. When you click on them, the next page of the experiment will appear.

Part I

In the first part of the experiment, we ask you to truthfully fill out a questionnaire. This is a personality questionnaire, so there are no right or wrong answers.
You receive 2 € for the completion of the questionnaire.
Please answer the questions with the answer options:
• Does not apply • Tends not to apply • Tends to apply • Applies
[The questionnaire items are shown in Section C.1.2.]

Part II

Screen 1: Explanation Ranking

In the further course of the experiment, you will simultaneously interact with **one** other participant (your *team partner*). This **interaction will take about 5-10 minutes**.

Important: For that part of the experiment you will receive a payoff of 2 €.

There are **4 possible partners for the interaction.**

In the following, you will be able to indicate who you would prefer as a team partner. For this you will see a randomly selected part of your potential partners' answers from the questionnaire.

At the same time, your potential partners will see your answers (as well as the answers of other participants) to these questions.

After viewing the profiles, we ask you to **indicate a preference order**.

With this preference order **you indicate with whom of the participants you would prefer to interact**. Rank 1 means that you would most like to have this participant as your team partner. Rank 2 means that you would second most like to have this participant as your team partner, and so on.

Important: You maximize your chances to interact with your preferred partner by indicating your true preferences.

This is intuitively the case, since the allocation mechanism

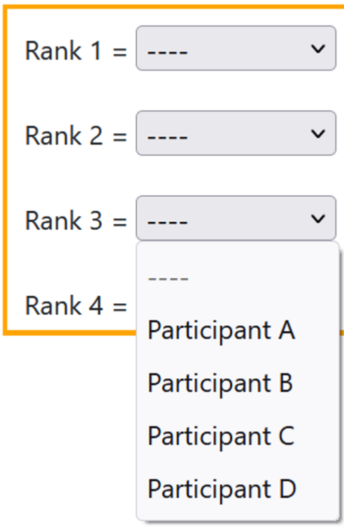
1. sorts all participants randomly,
2. gives the first participant his preferred team partner,
3. gives the second participant his preferred team partner from the remaining pool of possible partners,
4. and continues with this process until all participants have a partner.

Screen 2: Choice of Team Partner

Sample screen:
Below you will find information from Part I about your potential team partners.

	Participant A	Participant B	Participant C	Participant D
Statement 1: I support quota regulations in the labor market for socially disadvantaged groups (e.g., for women or migrants).	Applies	Tends to apply	Tends not to apply	Tends not to apply
Statement 2: There should be a requirement to wear a bicycle helmet.	Does not apply	Applies	Does not apply	Does not apply
Statement 3: I would rather spend an evening in a bar than partying in a club.	Applies	Tends to apply	Does not apply	Applies
Statement 4: I am easygoing, prone to laziness.	Tends to apply	Tends to apply	Tends to apply	Does not apply
Statement 5: I prefer to live in a shared apartment than alone.	Does not apply	Does not apply	Tends to apply	Tends to apply

We now ask you to specify a preference order. With this preference order you indicate with which of the participants you would prefer to interact.



Information after the Choice of the Team Partner

Thank you for submitting your preferences. We will inform you later which participant you will work with in the interaction task in Part IV.

Before that, however, you will make some decisions and assessments in Part III.

Part III

Screen 1: Fermi-Problem, Instructions and Comprehension Questions

In Part III, you will be confronted with a problem that is also known as Fermi-Problem:

A Fermi-Problem is a **quantitative estimate** for a **problem** for which initially **little data is available**.

Therefore, you will **hardly be able to answer such a problem accurately at first go**. However, after **careful and structured thought**, you will be able to **converge to the correct answer**. This is called an **educated guess**.

The classic example of a Fermi-Problem is the question about the **number of piano tuners in Chicago**.

It is obvious that you do not know the exact solution to this question.

The best strategy is to **break the problem into parts** that can be worked on. This means that you need to make some assumptions in order to start a calculation based on them.

In this example, the following assumptions might be useful:

- Number of Chicago residents.
- Percentage of households with a piano.
- Frequency with which a piano is tuned.
- Duration to tune a piano.
- Weekly working hours of a piano tuner.

It is **not always easy to estimate the exact value at each step**.

It is often **easier and more reliable to estimate the upper and lower limits than the direct value**. How does one arrive at the best estimate based on the upper and lower limits?

It turns out that the average of upper and lower bounds is not the best estimate. The **geometric mean is a much better choice**. The geometric mean of two numbers is the square root of their product. For example, the geometric mean of 5 and 20 is 10, because $10 = \sqrt{5 * 20}$. Thus, if you are sure that Chicago's population is at least 1 million and at most 9 million, it makes sense to take $\sqrt{1 * 9} = 3$ million as an estimate. You can use the calcalator that is provided to you on the desk.

It is also often helpful to look at a problem from two different angles. Your confidence in your educated guess will increase if you arrive at a similar number another way.

Your **payoff will depend on the accuracy of your answer**. More detailed information on this is provided hereafter.

You have **limited time to solve the problem**. This will be shown to you on the screen.

Comprehension Questions

A Fermi problem refers to a question that many people can spontaneously answer correctly using their factual knowledge?

❍ Yes

❍ No

A Fermi problem is best solved by breaking it down into parts?

❍ Yes

❍ No

It makes sense to determine upper and lower limits for your estimates?

❍ Yes

❍ No

The arithmetic mean is preferable to the geometric mean in estimation?

❍ Yes

❍ No

Screen 2: Individual Fermi-Problem, Instructions

You will now be given a **first Fermi-Problem to solve on your own**.

You can **receive a maximum of 5 euros** for solving this problem.

- Your payoff is calculated as follows:
    - Your answer deviates by **less than 10% from the true value: 5 Euro**
    - Your answer deviates by **less than 20% from the true value: 4 Euro**
    - Your answer deviates by **less than 40% from the true value: 3 Euro**
    - Your answer deviates by **less than 60% from the true value: 2 Euro**
    - Your answer deviates by **less than 80% from the true value: 1 Euro**
    - Your answer deviates by **more than 80% from the true value: 0 Euro**

Time for solving the problem: 5 minutes

Screen 3: Individual Fermi-Problem

How many weddings were performed in Germany in July 2022?

[Info: You can drag the text field larger with the mouse at the bottom right corner.]

Logical Steps:

1. 2. 3. 4. 5. ...

Submit final response

Screen 4: Joint Fermi Problem, Instructions and Comprehension Questions

Explanation

In the following, you will solve **another Fermi-Problem**. This time you will work **together with a team partner**.

You have **8 minutes to solve it**. You can make **up to 10 euros**.

Important: Your answer only counts if **both team partners give the identical answer**.

Your team partner is **one of the four participants whose answers you have seen and based on which you have given a preference order for the interaction task**.

For the Fermi-problem, you were **randomly assigned one of the four participants**.

On the next page, you will learn who your team partner is for the Fermi-Problem.

In doing so, you will learn:

1. **How you ranked your team partner.**
2. ***How your team partner ranked you. [only in the treatment info]***
3. **What answers your team partner gave.**

Comprehension Questions

You have to enter exactly the same result as your team partner for it to count?

❍ Yes

❍ No

You have been randomly assigned your team partner?

❍ Yes

❍ No

You are about to find out how your team partner ranked you?

❍ Yes

❍ No

Screen 5: Joint Fermi Problem, Your Team Partner

Your submitted preference order:

Rank 1: Participant A
Rank 2: Participant C
Rank 3: Participant B
Rank 4: Participant D

- **Your assigned partner is: Participant C**

Below you will find the information from Part I about your team partner.

Information about your team partner

Continue

Screen 6: Joint Fermi Problem, Assessments

Before you perform the task with your team partner, **we will ask you for a few more assessments on the upcoming teamwork.**

You can earn extra money from two of the assessments. We will randomly select one of the two assessments and pay you the amount earned. You will get more detailed information about these decisions on the next pages.

With other decisions you can influence the course of the further experiment.

Screen 7: Joint Fermi Problem, Assessment 1: Your Motivation

Here we ask you for a subjective assessment of how motivated you are to make an effort in the following task with your team partner.

How motivated are you to make an effort in the following task (from 0 to 10)?

Screen 8: Joint Fermi Problem, Assessment 2: Motivation of your partner

Should you guess your team partner's answer to the previous question correctly, you will receive 2 €. If you estimate the information incorrectly, you will receive 0 €.

How much do you think your partner is motivated to do the following task (from 0 to 10)?

Screen 9: Joint Fermi Problem, Assessment 3: Success of your team

Now we ask you for an **estimation of the success of your team**. Again, you can receive money for a correct estimation.
For this estimation you have a **credit of 1 €, with which you can bet on the performance of your team**. You can **decide which part of the €1 you want to keep and which part you want to bet**.
The part you bet

- is multiplied by 4 if your team is in the top 25% of today's teams.
- is lost if your team is not in the top 25% of today's teams.

We illustrate this with two examples.

1. If you bet 1 € and your team is in the top 25% of today's teams, you receive 4 €. On the other hand, if you bet 1 € and your team is not in the top 25%, you receive 0 €.
2. if you bet 0,50 € and your team is among the best 25%, you receive 2,50 € (0,50 €*4 + 0,50 €). If you bet 0,50 € and your team is not in the top 25%, you receive 0,50 €.

How much do you want to bet (0 - 1 €)?

Screen 10: Joint Fermi Problem, Assessment 4: Alone or with partner in additional Fermi-Problem with fixed payoff?

It is **possible that in the further course of the experiment you will have to solve an additional Fermi-Problem** where you will **not be paid for the accuracy of your data**.

Instead, you will receive a **fixed payoff for this Fermi-Problem** that is not related to your performance.

The only thing we will check is whether you work conscientiously on the Fermi-Problem. The duration of the task will again be 5 minutes.

You can indicate whether you prefer to do this **task alone or with your current team partner**.

This indication is based on 11 choices, each of which requires you to choose Option A or Option B. Each of the rows in the table below represents one decision.

- At the end of the experiment, one of the 11 decisions is selected. The computer will randomly select (with equal probability) which decision will be chosen.
- If either you or your team partner chose Option A in this decision, you will work the Fermi-Problem alone.
- If you both chose Option B, a random mechanism will decide whether you will do the task individually or together again. Thus, your team partner will never know which decision you made (and you will never know which decision your team partner made).

Option A			**Option B**
Do the task alone and get 1.0€	❍	❍	Do the task together and get 2.0€
Do the task alone and get 1.1€	❍	❍	Do the task together and get 1.9€
Do the task alone and get 1.2€	❍	❍	Do the task together and get 1.8€
Do the task alone and get 1.3€	❍	❍	Do the task together and get 1.7€
Do the task alone and get 1.4€	❍	❍	Do the task together and get 1.6€
Do the task alone and get 1.5€	❍	❍	Do the task together and get 1.5€
Do the task alone and get 1.6€	❍	❍	Do the task together and get 1.4€
Do the task alone and get 1.7€	❍	❍	Do the task together and get 1.3€
Do the task alone and get 1.8€	❍	❍	Do the task together and get 1.2€
Do the task alone and get 1.9€	❍	❍	Do the task together and get 1.1€
Do the task alone and get 2.0€	❍	❍	Do the task together and get 1.0€

Screen 10: Joint Fermi Problem

Sample screen:

Remaining time: 4:48

Instructions

Your team partner (Information)

How many dogs are there in Germany?

Please indicate your **answer in millions** (as decimal character you can use both “," or “.").

Chatbox: Please communicate here with you team partner.

Submit final response

Screen 11: Joint Fermi Problem, Cooperation Questionnaire

Please now indicate how much the following statements are true in relation to working with your team partner.
Please answer the questions with the answer options:
• Do not agree at all • Do not agree • Neutral • Agree • Agree fully
[The questionnaire items are shown in Section C.1.4.]

Screen 12: Joint Fermi-Problem, Alone or with team partner in additional Fermi-Problem?

It is **possible that you will perform another round of the identical task later in the experiment**, where you will again be **paid for the accuracy of your data**.

You can **specify whether you prefer to perform this task alone or with your current team partner**.

- If either you or your team partner selected "Alone", you will work the Fermi-Problem alone.
- If you both chose "Team Partner", a random mechanism will decide whether you do the task individually or together again. Thus, your team partner will never know what decision you made (and you will never know what decision your team partner made).

Would you prefer to do another round of the task alone, or with your team partner?
❍ Alone
❍ Team partner

Part IV

Screen 1: Interaction Task, Instructions

- Your task is to create **3 questions** for the game **“I have never ...**".
- The game works as follows:
  - A player starts a sentence with "I have never..." and ends it with something he or she has never done before.
  - If one of the other players has done that thing before, he or she gets a point deducted.
- Please create **statements where you think about half of the players have done this thing**, but half have not.
- Think other players as “average Munich students".
- Important: Again, both team partners have to enter the same answers. However, it is **not** crucial whether this is absolutely identical in wording

Your team partner
(based on the allocation mechanism)

Information

Screen 2: Interaction Task

[not displayed]

Part V

Screen 1: Additional task, Revelation

Chance has decided. You

1. do not solve an additional Fermi-Problem.
2. solve an additional Fermi-Problem with a fixed payment.
3. solve an additional Fermi-Problem with a performance-contingent payment.

Screen 2 (possible): Additional task

[not displayed]

Part VI

Screen 1: Socio-Demographic Questionnaire

Please provide the following statistical information.

- Age [integer]
- Gender [male; female; diverse]
- Field of study (faculty/major) [string]
- What language(s) is (are) your native language(s)? [string]
- What is your high school graduation grade? [number; 1-6]
- What is your high school graduation grade in mathematics? [number; 1-6]
- How many times have you participated in an economic laboratory study (including outside of this laboratory)? [0; 1-2; 3-5; 5+]

D

Appendix to Chapter 4

Gendered Access to Entrepreneurial Finance

D.1 Introduction to the Study and Informed Consent

In this section, we provide an overview of the introductory session that facilitators delivered to participants prior to the actual study. After an introduction from the branch manager, the facilitator outlined the study's objectives: Participants would be evaluating five business ideas and expressing their opinions. The focus of the study on gender was never mentioned.

We familiarized the participants with the questionnaire's structure using flip charts. We visually illustrated the survey procedures and the structure of incentives as outlined in detail in Section D.2. For the more complex procedure of the investment decision, we also used a graphical depiction of both states of the world to make sure the incentive scheme is understandable. We also illustrated any technical details such as how to operate a slider for any participants previously unfamiliar with such survey response techniques. Participants also had a chance to post any clarification questions.

Investment Graphical Depiction

Finally, the facilitator explicitly sought participants' informed consent. Participants were informed that their participation is voluntary, and that they have the right to withdraw at any point without facing any negative consequences.

Following the introduction, participants proceeded to their workstations to begin the survey. The facilitators were present at all times to resolve any comprehension or technical issues during the study.

D.2 Survey

In this section, we present a summary of the survey's structure. The survey was administered using the software Qualtrics.

Instructions and Comprehension Check

When siting at their workstation, participants first encounter two comprehension questions designed to evaluate their understanding of the incentivized investment questions. These questions are crucial to ensure that participants understood the investment question presented to them in the introduction of the study. The comprehension questions are as follows:

1. *Imagine that you invest 2,500 UGX to the business and keep 2,500 UGX. The business reported that it still exists and makes profits. How much do you have in total?*
2. *Imagine that you invest 4,000 UGX to the business and keep 1,000 UGX. The business reported that it does not exist anymore. How much will you have in total?*

The participants can proceed with the survey once they have correctly answered the two comprehension questions.

Pitch Deck Presentation

Participants are introduced to the name of the business idea and the identity of the **founder** associated with it. The name of the founder will vary randomly, while all other information remains constant across participants.

Green Market is a business idea originated by **{Name}** during the entrepreneurship academy.

{Name} is 24 years old and has a bachelor's degree in business administration and information technologies. **{Name}** did an internship at an agribusiness for six months.

In the next pages, you will able to see the idea of **{Name}** presented at the end of the entrepreneurship academy.

Take your time to go over **{Name}** 's idea.

Founder Description

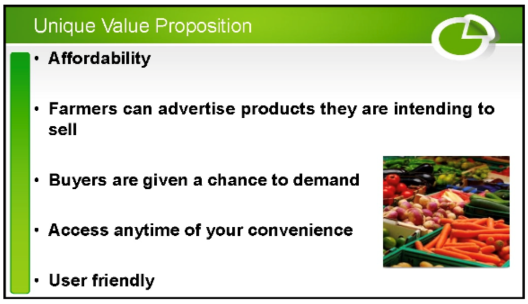
Following the founder's description, participants are presented with the pitch deck detailing the business idea. Below is the first business idea, *Green Market*.

D. APPENDIX TO CHAPTER 4

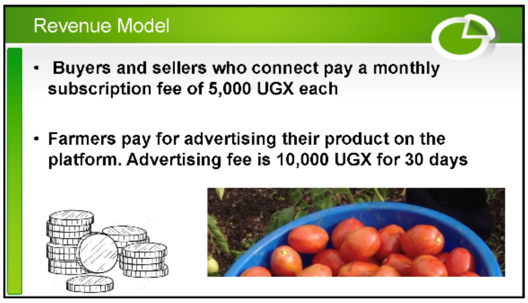
1

2

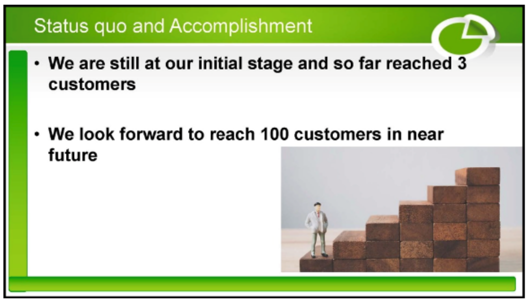
3



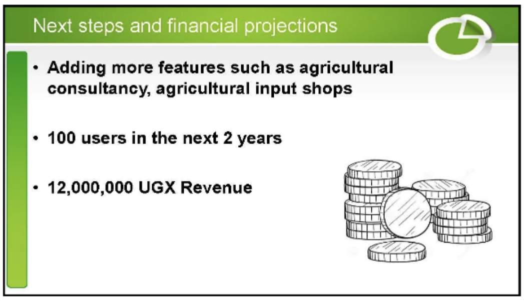
4



5



6



7

8

After reviewing the business idea, participants are presented with the following two questions.

Please rate your level of agreement from 0-100 with the following statements:

1. *This business idea meets a need or solves a problem in Uganda.*
2. *There is a market for this business idea in Uganda.*

In the subsequent step, participants receive information about the **implementer** of the idea, presented in a similar manner as the founder's description. For instance, for the business idea *Green Market*, the implementer's information reads as follows:

The candidate to implement this idea is {Name}. {Name} is 25 years old and holds a degree in business administration and information technologies. {Name} also participated in entrepreneurship academies and completed a semester-long internship at a farming enterprise.

Upon reviewing the implementer's information, participants proceed to the questions related to the pitch deck.

Attention Check

Before proceeding to the main survey questions, participants are required to answer a question designed to ensure their attentiveness to the survey's content. The question includes three potential answers, with only one of them being correct. This verifies that the participants actually pay attention and process the information about the gender of the founder and the implementer.

Before we do that, we just want to check what you remember about the description of Green Market. Which statement is true?

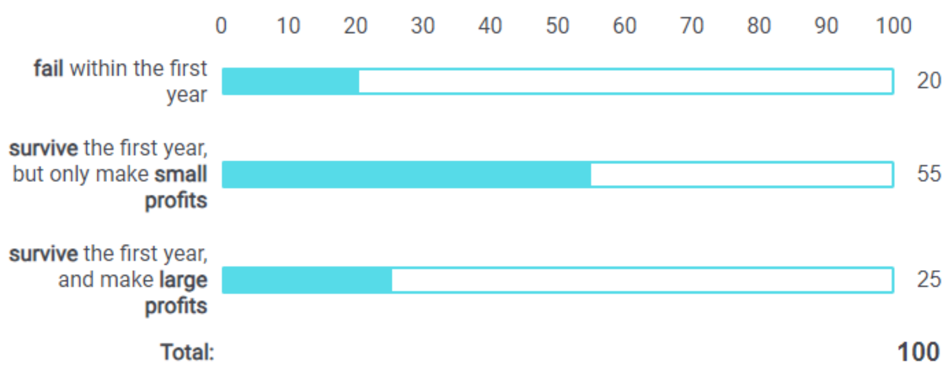
Business idea originated by: **{Name}**
Implemented by: **{Name}**

- {Name} came up with the idea of Green Market during an entrepreneurship academy.
- Green Market was not implemented by {Name}
- Green market is a company that focuses on production of renewable electricity.

Probabilistic Beliefs about Business Success

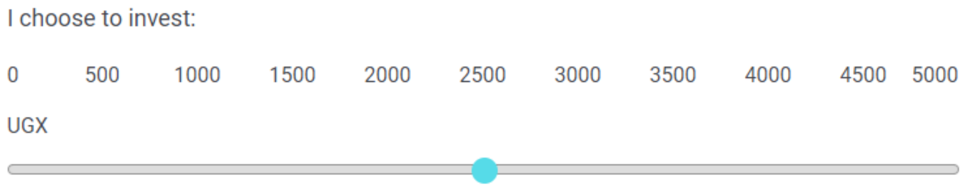
The first set of questions asks participants to estimate the likelihood of three different types of business outcomes.

What is the chance that this business idea will...



Investment

In this question, participants determine the amount they want to invest in the business, ranging from 0 to 5,000 UGX. They see a slider on the screen to facilitate their decision-making process.



In total, participants see five pitch decks. This means that the sequence from "Pitch Deck Presentation" to this point is repeated five times, once for each pitch deck.

Information Request

Participants can revise the amount they initially decided to invest in the business that was selected to be payoff-relevant. This comes as a surprise to them. For this, participants can request additional information. They are informed about this in the following way:

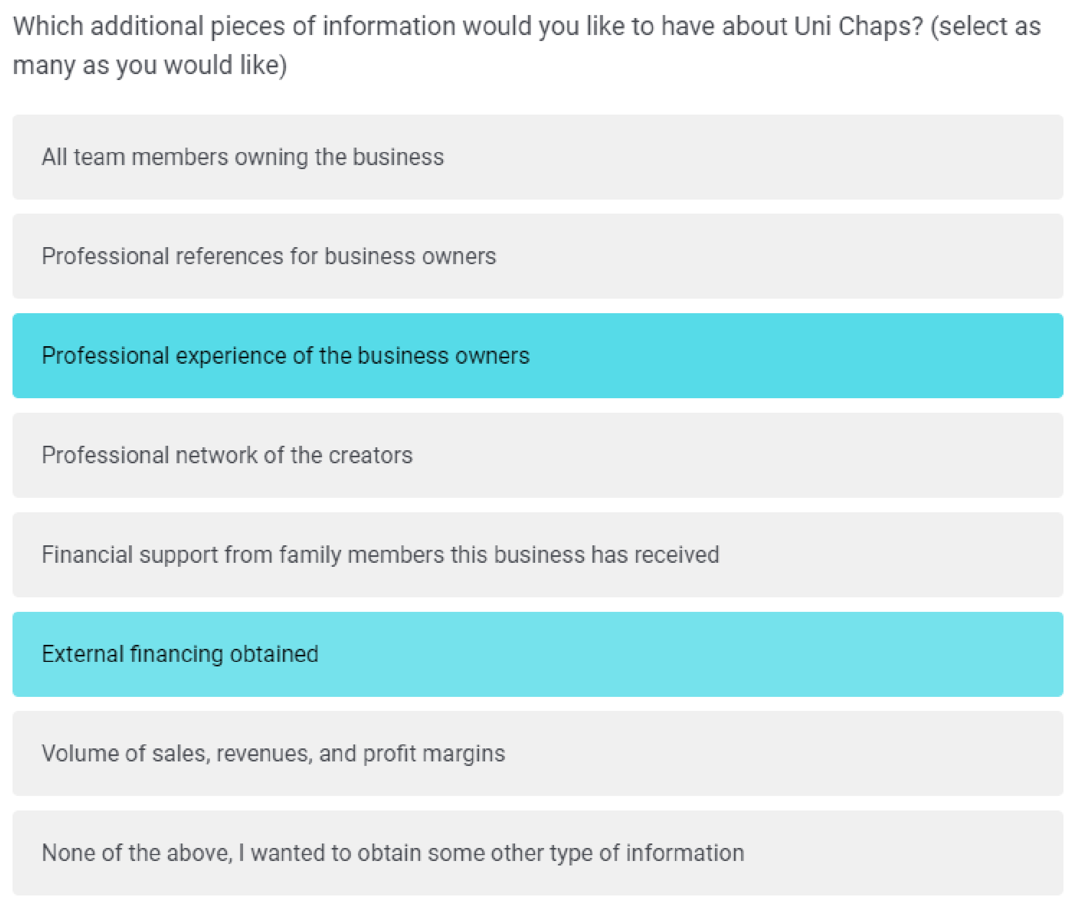
Uni Chaps was chosen as the business for which the investment decision will determine the potential bonus. You previously decided to invest {amount} UGX in this business. We would like to offer you the chance to revise your investment. To make an informed decision, you have the option to obtain additional information about Uni Chaps. However, it is essential to note that acquiring and verifying this information incurs a cost. Just as in your job, gathering and assessing information about a borrower's business entails costs, such as the time spent on this task.

1. *Yes, I want to check which additional information is available.*
2. *No, I will revise my investment decision without additional information.*

If the participant selects 'Yes', the available information and its associated cost are explained as follows: *Each piece of additional information, if available, costs 200 UGX.*

If you choose to access a particular piece of information, we will deduct 200 UGX from your final payoff. Conversely, if the requested information is not available, there will be no charge, and your payoff will not be affected.

Participants are provided with a clear understanding of the cost implication associated with acquiring additional information, enabling them to make an informed decision on whether to proceed with obtaining the information or not.


Which additional pieces of information would you like to have about Uni Chaps? (select as many as you would like)

- All team members owning the business
- Professional references for business owners
- Professional experience of the business owners
- Professional network of the creators
- Financial support from family members this business has received
- External financing obtained
- Volume of sales, revenues, and profit margins
- None of the above, I wanted to obtain some other type of information

If the participant selects 'No', they have the opportunity to revise their investment without seeing which pieces of information may be available.

Best Performing Business

In this question, participants select the business that they think is the most profitable of the five businesses they have encountered Participants are incentivized to make an accurate prediction, as a correct choice will result in a bonus reward:

Which business do you think was generating the highest profits when we last contacted them? If you guess correctly, you will earn an additional 5,000 UGX bonus. However, if your guess is incorrect, you will not receive any bonus.

Gender Norms

Participants are asked to indicate their level of agreement with the following statements using a 5-point Likert scale:

(a) A man's job is to earn money; a woman's job is to look after the home and family.

(b) Family life suffers when the woman has a full-time job.

(c) A job is all right, but what most women really want is a home and children.

Demographics

The survey concludes with three demographic questions:

1. *How old are you?*
2. *What is your gender? (Male / Female / Other)*
3. *How long have you been a loan officer?*

D.3 Tables

Table D.1: Founder and Implementer Names

Male	**Female**
Benjamin	Alinda
David	Carolin
Derrick	Dorothy
Duncan	Elisabeth
Kelvin	Esther
Martin	Juliana
Nicholus	Olivia
Ivan	Patience
Joel	Rebecca
Richard	Vanessa

Notes. Names used to signal the gender of founder and implementer.

Table D.2: Balance Test

	Male Entrepreneur		Female Entrepreneur			
	Mean (1)	SD (2)	Mean (3)	SD (4)	Difference (5)	p-value (6)
Age	34.0	5.04	34.2	5.24	-0.15	0.75
Male	0.57	0.50	0.54	0.50	0.03	0.54
Experience	6.84	4.06	6.73	4.35	0.12	0.77

Notes. Comparison of loan officer characteristics who either saw the individual entrepreneur as being male or female. Column (1)-(4) show the mean and the standard deviation of the characteristics across both sample, Column (5) reports the difference between the means, Column (6) shows the p-value of a two-sample t-test.

Table D.3: Investment and Best Business Decision with Individual Controls [Individual Entrepreneurs]

Panel A: Investment		Gender bias		LO gender		Experience	
		Low	High	Female	Male	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Entrepreneur	-232.98**	-125.77	-384.55**	-284.84	-219.13	-270.68	-214.39
	(118.50)	(171.93)	(169.40)	(179.38)	(159.43)	(184.55)	(159.83)
Age	-7.57	1.63	-14.75	-40.78	13.12	27.47	-50.00**
	(17.12)	(22.47)	(26.62)	(25.64)	(21.52)	(22.92)	(23.11)
Experience	31.09	13.50	55.64*	66.88**	10.43	-20.91	88.01***
	(19.64)	(26.29)	(30.00)	(30.04)	(24.18)	(57.79)	(28.44)
Male	268.88**	140.92	292.43*			254.86	273.40*
	(120.70)	(179.65)	(162.14)			(178.97)	(164.08)
Mean Dep. Var.	3,490.7	3,342.1	3,656.9	3,333.3	3,609.8	3,442.7	3,529.2
Observations	451	234	217	201	250	199	252
Panel B: Best Business							
Female Entrepreneur	-0.07*	-0.04	-0.10*	-0.11*	-0.04	-0.14**	-0.04
	(0.04)	(0.05)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
Age	0.00	0.01	-0.00	-0.01	0.01	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Experience	-0.01**	-0.02*	-0.01	0.00	-0.03***	-0.02	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Male	0.07*	0.01	0.11*			0.12*	0.05
	(0.04)	(0.06)	(0.05)			(0.06)	(0.05)
Mean Dep. Var.	0.26	0.25	0.27	0.24	0.28	0.33	0.21
Observations	451	234	217	201	250	199	252

Notes. OLS Regressions of the dependent variable *Investment* (Panel A) or *Best Business* (Panel B) on the gender of the individual entrepreneur who founded and implemented the business. Panel A reports the incentivized decision of how much to invest in the pitch deck business from 0-5,000 UGX. Panel B reports the incentivized decision of which of all the businesses is the best idea of all the ones seen; it is a probability. Column (1) reports the average effect, and Columns (2)-(7) split the observations according to different relevant observable characteristics. Columns (2)-(3) split by gender bias following International Social Survey Programme gender bias metrics. It incorporates three questions: *A man's job is to earn money; a woman's job is to look after the home and family, a job is alright, but what most women really want is home and children,* and *family life suffers when the woman has a full-time job*. The sample is split at the median. (4)-(5) are split according to the self-reported gender of the respondent, and (6)-(7) are split according to the median experience level. We control for the loan officers' age, their years of experience in the bank, as well as their gender. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust and reported in parentheses. The table includes pitch deck FE in both Panels. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.4: Attentive Participants [Individual Entrepreneurs]

	Investment (1)	Best Business (2)	Request Info (3)	# Info (4)	P[failure] (5)	P[small profits] (6)	P[large profits] (7)
Female Entrepreneur	-219.40*	-.06	.06	.02	2.98	-2.95	-.04
	(123.53)	(.04)	(.05)	(.14)	(2.32)	(2.28)	(2.69)
Mean Dep. Var.	3461	.25	.36	.69	24	41	35
Observations	410	410	410	410	410	410	410

Notes. OLS Regressions. The sample of attentive participants excludes the choices of ten percent of individuals who clicked most on the survey page when answering comprehension questions. The dependent variable is the gender of the individual entrepreneur who founded and implemented the business. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. The table includes round FE except for Columns (3) & (4). Standard errors are heteroskedasticity-robust and reported in parentheses. *** p <0.01, ** p <0.05, * p <0.1.

Table D.5: Beliefs about Business Success [Individual Entrepreneurs]

	P[failure] (1)	P[small profits] (2)	P[large profits] (3)
Female Entrepreneur	4.081*	-2.858	-1.223
	(2.187)	(2.210)	(2.568)
Mean Dep. Var.	22.47	41.80	35.73
Observations	451	451	451

Notes. OLS Regressions of the probability of the realization of three different scenarios about the business success on the gender of the individual entrepreneur who founded and implemented the business. 100 points could be allocated among the three different scenarios. The question was phrased as follows: *What is the chance that this business idea will 1) fail within the first year, 2) survive the first year, but only make small profits, and 3) survive the first year and make large profits*. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust and reported in parentheses. The table includes pitch deck FE in both Panels. *** p <0.01, ** p <0.05, * p <0.1.

Table D.6: Idea Quality [Individual Entrepreneurs]

	Idea Quality (0-100)				
	Round/Deck 1 (1)	Round/Deck 2 (2)	Round/Deck 3 (3)	Round/Deck 4 (4)	Round/Deck 5 (5)
Female Founder	.6	-1.8	-4.9	4.6	-6.2
	(4.1)	(5.2)	(4.8)	(4.0)	(3.9)
Mean Dep. Var.	69.12	63.30	71.05	74.05	75.85
Observations	86	81	92	85	107

Notes. OLS regressions of the perceived quality of the business ideas on the gender of the individual entrepreneur who founded and implemented the business. The index is based on two questions: *1) Does this business idea meet a need or solve a problem in Uganda?* and *2) Is there a market for this business idea in Uganda?* Participants rated their agreement on a scale ranging from 0 (*completely disagree*) to 100 (*completely agree*). *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.7: Correlations of Business Evaluation Measures [Individual Entrepreneurs]

	Investment (1)	Best Business (2)	Idea Quality (3)	P[failure] (4)	P[large profits] (5)
Investment	1.00				
Best Business	0.26 ***	1.00			
Idea Quality	0.52 ***	0.27 ***	1.00		
P[failure]	−0.38 ***	−0.16 ***	−0.46 ***	1.00	
P[large profits]	0.32 ***	0.20 ***	0.38 ***	−0.57 ***	1.00

Notes. Pearson correlation coefficients *** p <0.01, ** p <0.05, * p <0.1.

Table D.8: Investment and Best Business: Alternative Gender Bias Measure

Panel A: Individuals	Investment		Best Business	
	Low Bias (1)	High Bias (2)	Low Bias (3)	High Bias (4)
Female Entrepreneur	-111.78	-386.84**	-0.05	-0.10*
	(169.41)	(169.07)	(0.05)	(0.05)
Mean Dep. Var.	3,305.2	3,443.2	0.3	0.2
Observations	231	220	231	220
Panel B: Teams				
Female Founder	-133.02	145.95	-0.04	0.02
	(118.14)	(125.19)	(0.04)	(0.04)
Female Implementer	-141.99	23.03	-0.03	-0.01
	(126.86)	(107.47)	(0.04)	(0.04)
Female Founder&Implementer	142.53	-60.65	0.13**	0.04
	(178.27)	(160.70)	(0.06)	(0.06)
Mean Dep. Var.	3,238.64	3,398.30	0.19	0.19
Observations	924	880	924	880

Notes. OLS Regressions of the dependent variable *Investment* or *Best Business* on the gender of the individual entrepreneur who founded and implemented the business (Panel A) or on the gender of the founder and the implementer in teams in which these are different individuals (Panel B). *Investment* is the incentivized decision of how much to invest in the pitch deck business from 0-5,000 UGX. *Best Business* is the incentivized decision of which of all the businesses is the best idea of all the ones seen; it is a probability. The table splits the sample by gender bias following International Social Survey Programme gender bias metrics. The adjusted gender bias measure in this table only uses the following two questions questions: *A man's job is to earn money; a woman's job is to look after the home and family* and *a job is alright, but what most women really want is home and children*. The sample is split at the median. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust in Panel A, clustered at the individual level in Panel B, and reported in parentheses. The table includes pitch deck FE in both Panels, Panel B additionally includes individual FE. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.9: Investment Decision - Pre-Analysis Plan [Teams of Entrepreneurs]

Panel A: Investment		Gender bias		LO gender		Experience	
		Low	High	Female	Male	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female Founder	27.60	-109.48	174.52*	19.89	30.41	-16.01	67.55
	(65.54)	(94.83)	(90.36)	(91.32)	(92.29)	(96.56)	(89.15)
Female Implementer	-40.00	-30.97	-53.75	-88.76	-6.71	-68.52	-9.83
	(56.55)	(80.97)	(78.94)	(86.33)	(74.21)	(80.70)	(79.12)
Mean Dep. Var.	3,352.2	3,395.6	3,308.3	3,353.5	3,351.2	3,303.0	3,393.4
F-Statistic	.018	1.1	.92	.28	.033	.43	.2
P-value	.89	.31	.34	.6	.85	.51	.66
Observations	1804	936	868	804	1000	796	1008
Panel B: Best Business							
Female Founder	0.03	0.00	0.06**	0.04	0.03	0.04	0.03
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Female Implementer	0.02	0.02	0.01	-0.00	0.03	-0.02	0.05*
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mean Dep. Var.	0.21	0.23	0.18	0.27	0.16	0.21	0.21
F-Statistic	2.3	.24	2.4	.37	2.2	.16	2.8
P-value	.13	.62	.12	.54	.14	.69	.093
Observations	1804	936	868	804	1000	796	1008

*Notes.*OLS Regressions of the dependent variable *Investment* (Panel A) or *Best Business* (Panel B) on the gender of the founder and the implementer in teams in which these are different individuals. Panel A reports the incentivized decision of how much to invest in the pitch deck business from 0-5,000 UGX. Panel B reports the incentivized decision of which of all the businesses is the best idea of all the ones seen; it is a probability. Column (1) reports the average effect, and Columns (2)-(7) split the observations according to different relevant observable characteristics. (2)-(3) split by gender bias following International Social Survey Programme gender bias metrics. It incorporates three questions: *A man's job is to earn money; a woman's job is to look after the home and family, a job is alright, but what most women really want is home and children,* and *family life suffers when the woman has a full-time job*. The sample is split at the median. (4)-(5) are split according to the self-reported gender of the respondent, and (6)-(7) are split according to the median experience level. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are clustered at the individual level and reported in parentheses. The table includes pitch deck and individual FE in Panel A and pitch deck FE in Panel B. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.10: Correlations of Business Evaluation Measures [Teams of Entrepreneurs]

	Investment (1)	Best Business (2)	Idea Quality (3)	P[failure] (4)	P[large profits] (5)
Investment	1.00				
Best Business	0.18 ***	1.00			
Idea Quality	0.51 ***	0.21 ***	1.00		
P[failure]	−0.32 ***	−0.17 ***	−0.42 ***	1.00	
P[large profits]	0.33 ***	0.22 ***	0.37 ***	−0.61 ***	1.00

Notes. Pearson correlation coefficients *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.11: Investment Decision by Round [Teams of Entrepreneurs]

	Investment into Business (0-5,000 UGX)				
	Round/Deck 1 (1)	Round/Deck 2 (2)	Round/Deck 3 (3)	Round/Deck 4 (4)	Round/Deck 5 (5)
Female Founder	-209.6	-181.8	16.5	141.4	266.8
	(188.5)	(245.5)	(182.6)	(194.4)	(189.7)
Female Implementer	-257.7	-74.1	90.2	-116.3	64.2
	(177.1)	(230.2)	(181.1)	(191.2)	(184.8)
Female Founder&Implementer	228.8	3.5	62.4	141.2	-320.4
	(274.6)	(307.1)	(269.5)	(278.2)	(286.5)
Mean Dep. Var.	3,235.6	3,064.9	3,374.7	3,460.4	3,459.3
Indiv. FE	No	No	No	No	No
Indiv. Controls	No	No	No	Yes	No
Round FE	No	No	No	No	No
Observations	365	370	359	366	344

Notes. OLS Regressions of the five different investment decisions on the gender of the founder and the implementer in teams in which these are different individuals. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors are heteroskedasticity-robust and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table D.12: Attentive Participants [Teams of Entrepreneurs]

	Investment (1)	Best Business (2)	Request Info (3)	# Info (4)	P[failure] (5)	P[small profits] (6)	P[large profits]
Female Founder	1.62	-.02	.11	.27	-.01	1.87	-1.86
	(90.72)	(.03)	(.07)	(.24)	(1.64)	(1.58)	(1.81)
Female Implementer	-69.31	-.02	.03	-.13	-1.30	4.24***	-2.94
	(87.92)	(.03)	(.07)	(.19)	(1.53)	(1.57)	(1.88)
Female Founder&Implementer	53.40	.08*	-.03	.06	-.88	-4.14*	5.02*
	(127.08)	(.04)	(.11)	(.32)	(2.21)	(2.20)	(2.80)
Mean Dep. Var.	3321	.21	.4	.75	26	37	37
F-Statistic	.021	1.4	2	.65	1.6	1.2	.011
P-value	.89	.24	.16	.42	.2	.27	.92
Observations	1640	1640	332	332	1640	1640	1640

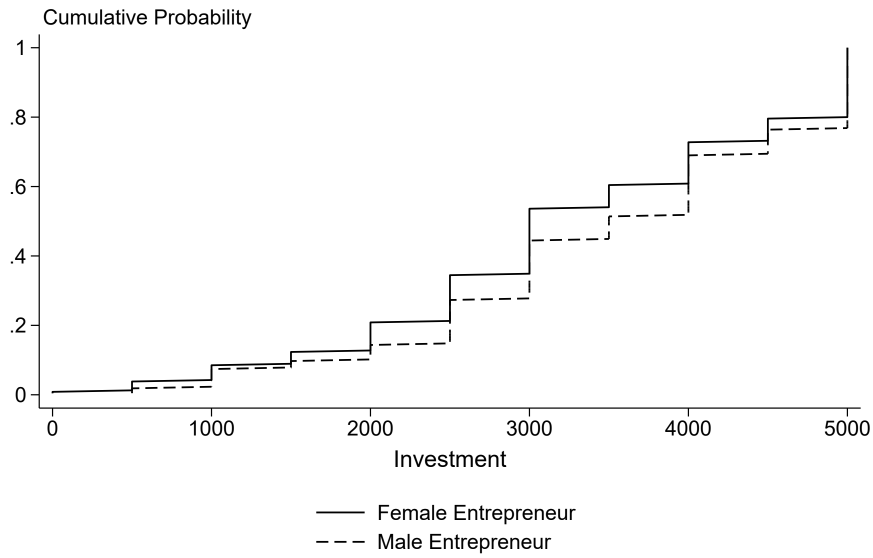
Notes. OLS Regressions. The sample of attentive participants excludes the choices of ten percent of individuals who clicked most on the survey page when answering comprehension questions. The dependent variable is the gender of the founder and the implementer in teams in which these are different individuals. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. The table includes round FE, individual FE and standard errors clustered at the individual level in Columns (1) & (5)-(7). Column (2) includes round FE and standard errors clustered at the individual level. Columns (2) & (3) do not include any FE, but heteroskedasticity-robust standard errors. *** p <0.01, ** p <0.05, * p <0.1.

Table D.13: Beliefs about Business Success [Teams of Entrepreneurs]

	P[failure] (1)	P[small profits] (2)	P[large profits] (3)
Female Founder	.240	.958	-1.198
	(1.394)	(1.377)	(1.633)
Female Implementer	-.398	2.084	-1.686
	(1.337)	(1.407)	(1.634)
Female Founder&Implementer	-1.572	-1.563	3.135
	(1.997)	(2.001)	(2.474)
Mean Dep. Var.	25.16	38.51	36.33
Observations	1804	1804	1804

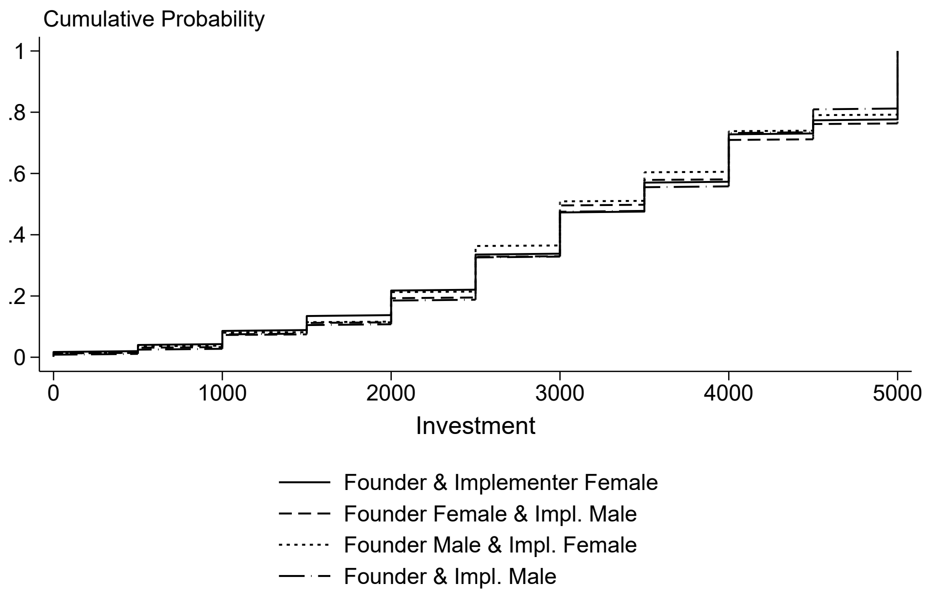
Notes. OLS Regressions of the probability of the realization of three different scenarios about the business success on the gender of the founder and the implementer in teams in which these are different individuals. 100 points could be allocated among the three different scenarios. The question was phrased as follows: *What is the chance that this business idea will 1) fail within the first year, 2) survive the first year, but only make small profits, and 3) survive the first year and make large profits*. *Mean Dep. Var* indicates the mean of the dependent variable of the reference group. Standard errors clustered at the individual level. The table includes round FE. *** p <0.01, ** p <0.05, * p <0.1.

D.4 Figures



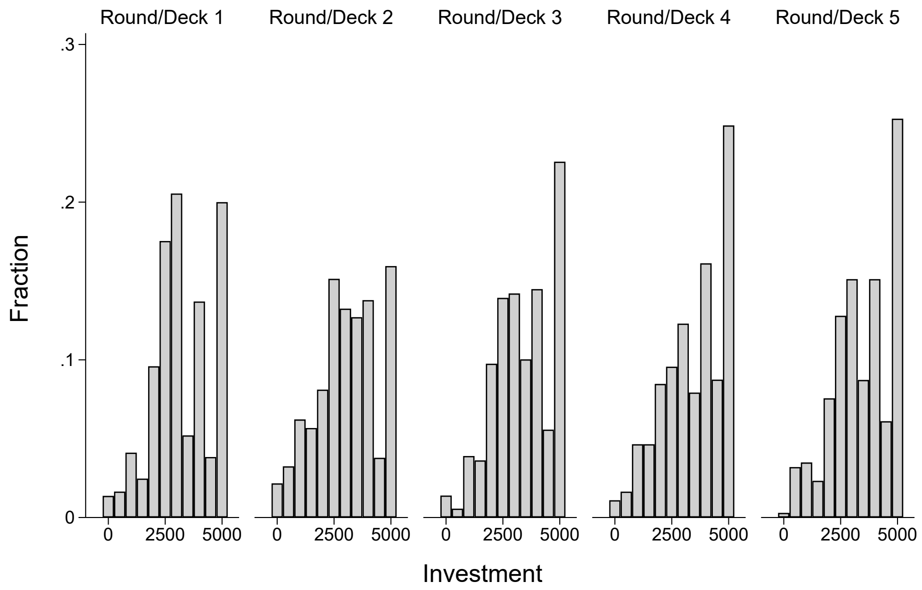
Notes. This figure the cumulative distribution functions of the invested amount into businesses of individual female or male entrepreneurs.

Figure D.1: Distribution Investment [Individual Entrepreneurs]



Notes. This figure the cumulative distribution functions of the invested amount into businesses of entrepreneurial teams with varying gender of the founder and implementer.

Figure D.2: Distribution Investment [Teams of Entrepreneurs]



Notes. This figure shows individual histograms for the invested amount into each of the pitch decks for the sample of entrepreneurial teams.

Figure D.3: Investment across Rounds/Pitch Decks [Teams of Entrepreneurs]

D.5 Deviations from the Pre-Analysis Plan

1. Our main specification was based on the assumption that there are no significant interaction effects between the gender of the founder and the gender of the implementer. Since this turned out not to be true (see e.g., Table 4.3), we rely on the secondary specification of the pre-analysis plan, which relaxed this assumption. However, we also present the results with the main specification of the PAP which only includes only dummies for the gender of the founder and implementer (but not the interaction), in Appendix Table D.9.

2. We do not split results by loan officer performance/productivity metrics because our partner was only able to provide these for about 50 percent of our sample (225 out of 451 loan officers).

List of Figures

List of Tables

Bibliography

Abdulkadiroğlu, A., & Sönmez, T. (1998). Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, *66*(3), 689–701.

Aerlebaeck, J. B., & Albarracin, L. (2019). The use and potential of fermi problems in the stem disciplines to support the development of twenty-first century competencies. *ZDM*, *51*(6), 979–990.

Agarwal, S., & Ben-David, I. (2018). Loan prospecting and the loss of soft information. *Journal of Financial Economics*, *129*(3), 608–628.

Agier, I., & Szafarz, A. (2013). Microfinance and Gender: Is There a Glass Ceiling on Loan Size? *World Development*, *42*(1), 165–181.

Aigner, D. J., & Cain, G. G. (1977). Statistical theories of discrimination in labor markets. *ILR Review*, *30*(2), 175–187.

Akerlof, G. A. (1982). Labor contracts as partial gift exchange. *Quarterly Journal of Economics*, *97*(4), 543–569.

Akerlof, G. A., & Kranton, R. E. (2000). Economics and identity. *Quarterly Journal of Economics*, *115*(3), 715–753.

Alesina, A. F., Lotti, F., & Mistrulli, P. E. (2013). Do women pay more for credit? Evidence from Italy. *Journal of the European Economic Association*, *11*, 45–66.

Alibhai, S., Donald, A. A., Goldstein, M., Oguz, A. A., Pankov, A., & Strobbe, F. (2019). Gender bias in SME lending: Experimental evidence from Turkey. *World Bank Policy Research Working Paper, No 9100*.

Allocca, A. (2023). 'No man is an island': An empirical study on team formation and performance. *CRC TRR 190 Discussion Paper No. 389*.

Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics*, *116*(1), 293–312.

Antler, Y. (2019). No one likes to be second choice. *Economic Journal*, *129*(619), 1119–1138.

Armstrong, M., & Zhou, J. (2016). Search deterrence. *Review of Economic Studies*, *83*(1), 26–57.

Aronson, E., & Worchel, P. (1966). Similarity versus liking as determinants of interpersonal attractiveness. *Psychonomic Science*, *5*(4), 157–158.

Arrow, K. (1973). Higher education as a filter. *Journal of Public Economics*, *2*(3), 193–216.

Ashraf, A. (2022). Performance ranks, conformity, and cooperation: Evidence from a sweater factory. *CESifo Working Paper No. 9591*.

Ashraf, N., Bandiera, O., Minni, V., & Quintas-Martınez, V. (2023). Gender roles and the misallocation of labour across countries. *Working Paper*.

Asiedu, E., Freeman, J. A., & Nti-Addae, A. (2012). Access to Credit by Small Businesses: How Relevant Are Race, Ethnicity, and Gender? *American Economic Review*, *102*(3), 532–37.

Åstebro, T., & Serrano, C. J. (2015). Business partners: Complementary assets, financing, and invention commercialization. *Journal of Economics & Management Strategy*, *24*(2), 228–252.

Autor, D., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, *118*(4), 1279–1333.

Autor, D., & Price, B. (2013). The changing task composition of the us labor market: An update of Autor, Levy, and Murnane (2003). *Working Paper*.

AVCA. (2022). *Venture Capital in Africa Report*. African Private Capital Association.

Avery, C., & Levin, J. (2010). Early admissions at selective colleges. *American Economic Review*, *100*(5), 2125–2156.

Bagues, M., & Esteve-Volart, B. (2010). Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *Review of Economic Studies*, *77*(4), 1301–1328.

Bandiera, O., Barankay, I., & Rasul, I. (2005). Social preferences and the response to incentives: Evidence from personnel data. *Quarterly Journal of Economics*, *120*(3), 917–962.

Bandiera, O., Barankay, I., & Rasul, I. (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association*, *11*(5), 1079–1114.

Banerjee, A. V., & Duflo, E. (2014). Do firms want to borrow more? Testing credit constraints using a directed lending program. *Review of Economic Studies*, *81*(2), 572–607.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. *Working Paper*.

Bartos, V., Czura, K., Kaiser, M., Opitz, T., & Shanks, B. (2023). Identifying and teaching high-growth entrepreneurship: Experimental evidence from entrepreneurship academies for university students in Uganda. *(Unpublished Manuscript)*.

Basteck, C., & Mantovani, M. (2018). Cognitive ability and games of school choice. *Games and Economic Behavior, 109,* 156–183.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117*(3), 497–529.

Beck, T., Behr, P., & Madestam, A. (2018). Sex and credit: Do gender interactions matter for credit market outcomes? *Journal of Banking & Finance*, *87*, 380–396.

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Bell, D. E. (1982). Regret in decision making under uncertainty. *Operations Research*, *30*(5), 961–981.

Bellucci, A., Borisov, A., & Zazzaro, A. (2010). Does gender matter in bank–firm relationships? Evidence from small business lending. *Journal of Banking & Finance*, *34*(12), 2968–2984.

Ben-Zur, H., & Breznitz, S. J. (1981). The effect of time pressure on risky choice behavior. *Acta Psychologica*, *47*(2), 89–104.

Berlin, N., & Dargnies, M.-P. (2016). Gender differences in reactions to feedback and willingness to compete. *Journal of Economic Behavior & Organization*, *130*, 320–336.

Bikhchandani, S., & Segal, U. (2014). Transitive regret over statistically independent lotteries. *Journal of Economic Theory*, *152*, 237–248.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. *Assessment and Teaching of 21st Century Skills*, 17–66.

Blanchflower, D. G., Levine, P. B., & Zimmerman, D. J. (2003). Discrimination in the small-business credit market. *Review of Economics and Statistics*, *85*(4), 930–943.

Bleichrodt, H., Cillo, A., & Diecidue, E. (2010). A quantitative measurement of regret theory. *Management Science*, *56*(1), 161–175.

Bohren, J. A., Haggag, K., Imas, A., & Pope, D. G. (2023). Inaccurate statistical discrimination: An identification problem. *Review of Economics and Statistics (Forthcoming)*.

Boss, V., Dahlander, L., Ihl, C., & Jayaraman, R. (2023). Organizing entrepreneurial teams: A field experiment on autonomy over choosing teams and ideas. *Organization Science*, *34*(6), 2097–2118.

Bowles, H. R., Babcock, L., & Lai, L. (2007). Social incentives for gender differences in the propensity to initiate negotiations: Sometimes it does hurt to ask. *Organizational Behavior and Human Decision Processes*, *103*(1), 84–103.

Bradler, C., Dur, R., Neckermann, S., & Non, A. (2016). Employee recognition and performance: A field experiment. *Management Science*, *62*(11), 3085–3099.

Braun, M., & Muermann, A. (2004). The impact of regret on the demand for insurance. *Journal of Risk and Insurance*, *71*(4), 737–767.

Brock, J. M., & De Haas, R. (2023). Discriminatory Lending: Evidence from Bankers in the Lab. *American Economic Journal: Applied Economics*, *15*(2), 31–68.

Brown, A. L., Imai, T., Vieider, F., & Camerer, C. (2023). Meta-analysis of empirical estimates of loss-aversion. *Journal of Economic Literature (Forthcoming)*.

Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences*, *106*(19), 7745–7750.

Buturak, G., & Evren, Ö. (2017). Choice overload and asymmetric regret. *Theoretical Economics*, *12*(3), 1029–1056.

BVerfG, 1 BvL 3/14, Urteil des Ersten Senats vom 19. Dezember 2017 (2017). https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2017/12/ls20171219_1bvl000314.html

Calderón, G., Iacovone, L., & Juarez, L. (2017). Opportunity versus necessity: Understanding the heterogeneity of female micro-entrepreneurs. *World Bank Economic Review*, *30*, S86–S96.

Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.-R., & Sirigu, A. (2004). The involvement of the orbitofrontal cortex in the experience of regret. *Science*, *304*(5674), 1167–1170.

Campos-Mercade, P. (2018). Power analysis through simulations in stata: A step by step guide.

Card, D., Cardoso, A. R., & Kline, P. (2015). Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women. *Quarterly Journal of Economics*, *131*(2), 633–686.

Carpenter, R. E., & Petersen, B. C. (2002). Is the growth of small firms constrained by internal finance? *Review of Economics and Statistics*, *84*(2), 298–309.

Chandrasekhar, A. G., Kinnan, C., & Larreguy, H. (2018). Social networks as contract enforcement: Evidence from a lab experiment in the field. *American Economic Journal: Applied Economics*, *10*(4), 43–78.

Charness, G., Cooper, D. J., & Grossman, Z. (2020). Silence is golden: Team problem solving and communication costs. *Experimental Economics*, *23*, 668–693.

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

Chen, R., & Gong, J. (2018). Can self selection create high-performing teams? *Journal of Economic Behavior & Organization*, *148*, 20–33.

Chen, Y., & Li, S. X. (2009). Group identity and social preferences. *American Economic Review*, *99*(1), 431–457.

Cohoon, J. M., Wadhwa, V., & Mitchell, L. (2010). *Are successful women entrepreneurs different from men?* Ewing Marion Kauffman Foundation.

Cole, S., Kanz, M., & Klapper, L. (2015). Incentivizing calculated risk-taking: Evidence from an experiment with commercial bank loan officers. *Journal of Finance*, *70*(2), 537–575.

Cooke, A. D., Meyvis, T., & Schwartz, A. (2001). Avoiding future regret in purchase-timing decisions. *Journal of Consumer Research*, *27*(4), 447–459.

Cooper, D. J., & Saral, K. J. (2020). The behavioral approach to entrepreneurship. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics* (pp. 1–25). Springer.

Coricelli, G., Critchley, H. D., Joffily, M., O'Doherty, J. P., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience*, *8*(9), 1255–1262.

Cox, J. C., & Oaxaca, R. L. (1989). Laboratory experiments with a finite-horizon job-search model. *Journal of Risk and Uncertainty*, *2*(3), 301–329.

Croson, R., & Gneezy, U. (2009). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448–74.

Currarini, S., Jackson, M. O., & Pin, P. (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, *77*(4), 1003–1045.

Currarini, S., & Mengel, F. (2016). Identity, homophily and in-group bias. *European Economic Review*, *90*, 40–55.

Davey, C. G., Allen, N. B., Harrison, B. J., Dwyer, D. B., & Yücel, M. (2010). Being liked activates primary reward and midline self-related brain regions. *Human Brain Mapping*, *31*(4), 660–668.

De Dreu, C. K., Evers, A., Beersma, B., Kluwer, E. S., & Nauta, A. (2001). A theory-based measure of conflict management strategies in the workplace. *Journal of Organizational Behavior*, *22*(6), 645–668.

Deming, D. J. (2017). The growing importance of social skills in the labor market. *Quarterly Journal of Economics*, *132*(4), 1593–1640.

Demirguc-Kunt, A., Klapper, L., Singer, D., & Ansar, S. (2018). *The Global Findex Database 2017: Measuring financial inclusion and the fintech revolution*. World Bank Publications.

Deversi, M., Kocher, M. G., & Schwieren, C. (2020). Cooperation in a company: A large-scale experiment. *CESifo Working Paper No. 8190*.

Dimotakis, N., Davison, R. B., & Hollenbeck, J. R. (2012). Team structure and regulatory focus: The impact of regulatory fit on team dynamic. *Journal of Applied Psychology*, *97*(2), 421.

Djankov, S., McLiesh, C., & Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, *84*(2), 299–329.

Dur, R. (2009). Gift exchange in the workplace: Money or attention? *Journal of the European Economic Association*, *7*(2-3), 550–560.

Dur, U., Pathak, P. A., Song, F., & Sönmez, T. (2022). Deduction dilemmas: The Taiwan assignment mechanism. *American Economic Journal: Microeconomics*, *14*(1), 164–185.

Eckel, C. C., & Grossman, P. J. (2005). Managing diversity by creating team identity. *Journal of Economic Behavior & Organization*, *58*(3), 371–392.

Economist. (2022). *African startups are raising unprecedented amounts. What next?*

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383.

Edmondson, A. (2002). *Managing the risk of learning: Psychological safety in work teams*. Division of Research, Harvard Business School Cambridge, MA.

Einav, L. (2005). Informational asymmetries and observational learning in search. *Journal of Risk and Uncertainty*, *30*(3), 241–259.

Engelbrecht-Wiggans, R., & Katok, E. (2008). Regret and feedback information in first-price sealed-bid auctions. *Management Science*, *54*(4), 808–819.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., & Schudy, S. (2021). The value of leadership: Evidence from a large-scale field experiment. *CESifo Working Paper No. 9273*.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., & Schudy, S. (2023a). The effect of incentives in non-routine analytical team tasks. *Journal of Political Economy (Forthcoming)*.

Englmaier, F., Grimm, S., Grothe, D., Schindler, D., & Schudy, S. (2023b). The efficacy of tournaments for non-routine team tasks. *Journal of Labor Economics (Forthcoming)*.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*(8), 709.

Ewens, M., & Townsend, R. R. (2020). Are early stage investors biased against women? *Journal of Financial Economics*, *135*(3), 653–677.

Fafchamps, M., & Woodruff, C. (2017). Identifying gazelles: Expert panels vs. surveys as a means to identify firms with rapid growth potential. *World Bank Economic Review*, *31*(3), 670–686.

Finucane, M. L., Alhakami, A., Slovic, P., & Johnson, S. M. (2000). The affect heuristic in judgments of risks and benefits. *Journal of Behavioral Decision Making*, *13*(1), 1–17.

Fioretti, M., Vostroknutov, A., & Coricelli, G. (2022). Dynamic regret avoidance. *American Economic Journal: Microeconomics*, *14*(1), 70–93.

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

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, *71*(3), 397–404.

Fischer, M., Rilke, R. M., & Yurtoglu, B. B. (2023). When, and why, do teams benefit from self-selection? *Experimental Economics*, 1–26.

Fisman, R., Paravisini, D., & Vig, V. (2017). Cultural proximity and loan outcomes. *American Economic Review*, *107*(2), 457–492.

Franke, N., Gruber, M., Harhoff, D., & Henkel, J. (2006). What you are is what you like—similarity biases in venture capitalists' evaluations of start-up teams. *Journal of Business Venturing*, *21*(6), 802–826.

Gächter, S., Johnson, E. J., & Herrmann, A. (2022). Individual-level loss aversion in riskless and risky choices. *Theory and Decision*, *92*(3), 599–624.

Gächter, S., Kölle, F., & Quercia, S. (2017). Reciprocity and the tragedies of maintaining and providing the commons. *Nature Human Behaviour*, *1*(9), 650–656.

Gächter, S., Starmer, C., & Tufano, F. (2022). Measuring "group cohesion" to reveal the power of social relationships in team production. *Review of Economics and Statistics (Forthcoming)*.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, *69*(1), 9–15.

Gerhards, L., & Kosfeld, M. (2020). I (don't) like you! But who cares? Gender differences in same-sex and mixed-sex teams. *Economic Journal*, *130*(627), 716–739.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.

Gill, D., & Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, *124*(6), 1619–1676.

Girotra, K., Terwiesch, C., & Ulrich, K. T. (2010). Idea generation and the quality of the best idea. *Management Science*, *56*(4), 591–605.

Goeree, J. K., McConnell, M. A., Mitchell, T., Tromp, T., & Yariv, L. (2010). The 1/d law of giving. *American Economic Journal: Microeconomics*, *2*(1), 183–203.

Goldin, C. (2020). Journey across a century of women. *NBER Reporter*, *3*, 1–7.

Gonzales Martinez, R., Aguilera-Lizarazu, G., Rojas-Hosse, A., & Aranda Blanco, P. (2020). The interaction effect of gender and ethnicity in loan approval: A Bayesian estimation with data from a laboratory field experiment. *Review of Development Economics*, *24*(3), 726–749.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125.

Guzman, J., & Kacperczyk, A. (2019). Gender gap in entrepreneurship. *Research Policy*, *48*(7), 1666–1680.

Hakimov, R., & Kübler, D. (2021). Experiments on centralized school choice and college admissions: A survey. *Experimental Economics*, *24*(2), 434–488.

Halpern, J. Y., & Leung, S. (2016). Minimizing regret in dynamic decision problems. *Theory and Decision*, *81*(1), 123–151.

Hardt, D., Mayer, L., & Rincke, J. (2023). Who does the talking here? The impact of gender composition on team interactions. *Working Paper*.

Hassidim, A., Romm, A., & Shorrer, R. I. (2021). The limits of incentives in economic matching procedures. *Management Science*, *67*(2), 951–963.

Hayashi, T. (2008). Regret aversion and opportunity dependence. *Journal of Economic Theory*, *139*(1), 242–268.

Hebert, C. (2020). Gender stereotypes and entrepreneur financing. *10th Miami Behavioral Finance Conference*.

Hedegaard, M. S., & Tyran, J.-R. (2018). The price of prejudice. *American Economic Journal: Applied Economics*, *10*(1), 40–63.

Hey, J. D. (1982). Search for rules for search. *Journal of Economic Behavior & Organization*, *3*(1), 65–81.

Hey, J. D. (1987). Still searching. *Journal of Economic Behavior & Organization*, *8*(1), 137–144.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655.

Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, *59*(7), 1514–1528.

Hoogendoorn, S., Parker, S. C., & Van Praag, M. (2017). Smart or diverse start-up teams? Evidence from a field experiment. *Organization Science*, *28*(6), 1010–1028.

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

Houser, D., & Winter, J. (2004). How do behavioral assumptions affect structural inference? Evidence from a laboratory experiment. *Journal of Business & Economic Statistics*, *22*(1), 64–79.

Hsieh, C.-T., & Olken, B. A. (2014). The missing “missing middle”. *Journal of Economic Perspectives*, *28*(3), 89–108.

Ibanez, M., Czermak, S., & Sutter, M. (2009). Searching for a better deal–on the influence of group decision making, time pressure and gender on search behavior. *Journal of Economic Psychology*, *30*(1), 1–10.

Ichniowski, C., & Shaw, K. (2013). Insider econometrics: A roadmap to estimating empirical models of organizational performance. In R. Gibbons & J. Roberts (Eds.), *The Handbook of Organizational Economics* (pp. 263–312). Princeton University Press.

IMF. (2021). *Financial Development Index Database*. International Monetary Fund.

Irons, B., & Hepburn, C. (2007). Regret theory and the tyranny of choice. *Economic Record*, *83*(261), 191–203.

Jain, P. (2020). Imperfect monitoring and informal insurance: The role of social ties. *Journal of Economic Behavior & Organization*, *180*, 241–256.

Jaoui, F., Amoussou, O., & Kemeze, F. H. (2022). “Catch me if you can” on drivers of venture capital investment in Africa. *African Development Review*, *34*, S117–S140.

Jhunjhunwala, T. (2021). Searching to avoid regret: An experimental evidence. *Journal of Economic Behavior & Organization*, *189*, 298–319.

Jin, G. Z., Jones, B., Lu, S. F., & Uzzi, B. (2019). The reverse Matthew effect: Consequences of retraction in scientific teams. *Review of Economics and Statistics*, *101*(3), 492–506.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*(9), 697.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kelley, D., Singer, S., & Herrington, M. (2016). *GEM 2015/2016 Global Report*. Global Entrepreneurship Monitor.

Kirchler, M., Andersson, D., Bonn, C., Johannesson, M., Sørensen, E. Ø., Stefan, M., Tinghög, G., & Västfjäll, D. (2017). The effect of fast and slow decisions on risk taking. *Journal of Risk and Uncertainty*, *54*(1), 37–59.

Kocher, M. G., Pahlke, J., & Trautmann, S. T. (2013). Tempus fugit: Time pressure in risky decisions. *Management Science*, *59*(10), 2380–2391.

Kocher, M. G., Schindler, D., Trautmann, S. T., & Xu, Y. (2019). Risk, time pressure, and selection effects. *Experimental Economics*, *22*(1), 216–246.

Kocher, M. G., & Sutter, M. (2006). Time is money—time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization*, *61*(3), 375–392.

Kogut, C. A. (1990). Consumer search behavior and sunk costs. *Journal of Economic Behavior & Organization*, *14*(3), 381–392.

Krähmer, D., & Stone, R. (2008). Regret in dynamic decision problems. *Working Paper*.

Kruglanski, A. W., & Freund, T. (1983). The freezing and unfreezing of lay-inferences: Effects on impressional primacy, ethnic stereotyping, and numerical anchoring. *Journal of Experimental Social Psychology*, *19*(5), 448–468.

Kube, S., Maréchal, M. A., & Puppe, C. (2012). The currency of reciprocity: Gift exchange in the workplace. *American Economic Review*, *102*(4), 1644–1662.

Kuziemko, I., Buell, R. W., Reich, T., & Norton, M. I. (2014). “Last-place aversion”: Evidence and redistributive implications. *Quarterly Journal of Economics*, *129*(1), 105–149.

La Porta, R., & Shleifer, A. (2008). The unofficial economy and economic development. *NBER Working Paper No. 14520*.

Lee, M., & Huang, L. (2018). Gender bias, social impact framing, and evaluation of entrepreneurial ventures. *Organization Science*, *29*(1), 1–16.

Lee, S., & Niederle, M. (2015). Propose with a rose? Signaling in internet dating markets. *Experimental Economics*, *18*(4), 731–755.

Leider, S., Möbius, M. M., Rosenblat, T., & Do, Q.-A. (2009). Directed altruism and enforced reciprocity in social networks. *Quarterly Journal of Economics*, *124*(4), 1815–1851.

Lemieux, J., & Peterson, R. A. (2011). Purchase deadline as a moderator of the effects of price uncertainty on search duration. *Journal of Economic Psychology*, *32*(1), 33–44.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, *1*(3), 593–622.

Lippman, S. A., & McCall, J. (1976). The economics of job search: A survey. *Economic Inquiry*, *14*(2), 155–189.

List, J. A. (2020). Non est disputandum de generalizability? A glimpse into the external validity trial. *NBER Working Paper No. 27535*.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*(368), 805–824.

Macchi, E. (2023). Worth your weight: Experimental evidence on the benefits of obesity in low-income countries. *American Economic Review*, *113*(9), 2287–2322.

Mascia, D. V., & Rossi, S. P. (2017). Is there a gender effect on the cost of bank financing? *Journal of Financial Stability*, *31*(100), 136–153.

McKenzie, D., & Sansone, D. (2019). Predicting entrepreneurial success is hard: Evidence from a business plan competition in Nigeria. *Journal of Development Economics*, *141*, 102369.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

McWhirter, R. M., & Jecker, J. D. (1967). Attitude similarity and inferred attraction. *Psychonomic Science*, *7*(6), 225–226.

Meisner, V., & von Wangenheim, J. (2023). Loss aversion in strategy-proof school-choice mechanisms. *Journal of Economic Theory*, *207*, 105588.

Michenaud, S., & Solnik, B. (2008). Applying regret theory to investment choices: Currency hedging decisions. *Journal of International Money and Finance*, *27*(5), 677–694.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making*, *5*(6), 437.

Möbius, M. M., Niederle, M., Niehaus, P., & Rosenblat, T. S. (2022). Managing self-confidence: Theory and experimental evidence. *Management Science*, *68*(11), 7793–7817.

Montoya, A. M., Parrado, E., Solis, A., & Undurraga, R. (2020). Bad taste: Gender discrimination in the consumer credit market. *IDB Working Paper No. 1053*.

Montoya, R. M., & Horton, R. S. (2012). The reciprocity of liking effect. In M. Paludi (Ed.), *The Psychology of Love* (pp. 39–57). Santa Barbara, CA: Praeger.

Montoya, R. M., & Horton, R. S. (2014). A two-dimensional model for the study of interpersonal attraction. *Personality and Social Psychology Review*, *18*(1), 59–86.

Montoya, R. M., Horton, R. S., & Kirchner, J. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. *Journal of Social and Personal Relationships*, *25*(6), 889–922.

Montoya, R. M., & Insko, C. A. (2008). Toward a more complete understanding of the reciprocity of liking effect. *European Journal of Social Psychology*, *38*(3), 477–498.

Moon, P., & Martin, A. (1990). Better heuristics for economic search—experimental and simulation evidence. *Journal of Behavioral Decision Making*, *3*(3), 175–193.

Morazzoni, M., & Sy, A. (2022). Female entrepreneurship, financial frictions and capital misallocation in the US. *Journal of Monetary Economics*, *129*, 93–118.

Morgan, J., Neckermann, S., & Sisak, D. (2021). Peer evaluation and team performance: An experiment on complex problem solving. *Working Paper*.

Muermann, A., & Volkman, J. M. (2007). Regret, pride, and the disposition effect. *Working Paper*.

Muravyev, A., Talavera, O., & Schäfer, D. (2009). Entrepreneurs' gender and financial constraints: Evidence from international data. *Journal of Comparative Economics*, *37*(2), 270–286.

Niederle, M. (2016). Handbook of Experimental Economics. *Princeton University Press*, 481–553.

OECD. (2017). *The Pursuit of Gender Equality-An Uphill Battle*.

Opitz, T., & Schwaiger, C. (2023a). Everyone likes to be liked: Experimental evidence from matching markets. *CRC TRR 190 Discussion Paper No. 366*.

Opitz, T., & Schwaiger, C. (2023b). Reciprocal preferences in matching markets. *Max Planck Institute for Innovation & Competition Research Paper No. 23-06*.

Oprea, R., Friedman, D., & Anderson, S. T. (2009). Learning to wait: A laboratory investigation. *Review of Economic Studies*, *76*(3), 1103–1124.

Pais, J., & Pintér, Á. (2008). School choice and information: An experimental study on matching mechanisms. *Games and Economic Behavior*, *64*(1), 303–328.

Pais, J., Pintér, Á., & Veszteg, R. F. (2011). College admissions and the role of information: An experimental study. *International Economic Review*, *52*(3), 713–737.

Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: Decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes*, *66*(2), 131–152.

Phelps, E. S. (1972). The statistical theory of racism and sexism. *American Economic Review*, *62*(4), 659–661.

Qin, J. (2015). A model of regret, investor behavior, and market turbulence. *Journal of Economic Theory*, *160*, 150–174.

Reb, J. (2008). Regret aversion and decision process quality: Effects of regret salience on decision process carefulness. *Organizational Behavior and Human Decision Processes*, *105*(2), 169–182.

Reb, J., & Connolly, T. (2009). Myopic regret avoidance: Feedback avoidance and learning in repeated decision making. *Organizational Behavior and Human Decision Processes*, *109*(2), 182–189.

Roth, A. E. (1991). Game theory as a part of empirical economics. *Economic Journal*, *101*(404), 107–114.

Salas, E., Reyes, D. L., & McDaniel, S. H. (2018). The science of teamwork: Progress, reflections, and the road ahead. *American Psychologist*, *73*(4), 593.

Sarver, T. (2008). Anticipating regret: Why fewer options may be better. *Econometrica*, *76*(2), 263–305.

Scholz, E., Jutz, R., Edlund, J., Öun, I., & Braun, M. (2014). ISSP 2012 family and changing gender roles IV: Questionnaire development.

Schotter, A., & Braunstein, Y. M. (1981). Economic search: An experimental study. *Economic Inquiry*, *19*(1), 1–25.

Schunk, D. (2009). Behavioral heterogeneity in dynamic search situations: Theory and experimental evidence. *Journal of Economic Dynamics and Control*, *33*(9), 1719–1738.

Schunk, D., & Winter, J. (2009). The relationship between risk attitudes and heuristics in search tasks: A laboratory experiment. *Journal of Economic Behavior & Organization*, *71*(2), 347–360.

Shastry, G. K., Shurchkov, O., & Xia, L. L. (2020). Luck or skill: How women and men react to noisy feedback. *Journal of Behavioral and Experimental Economics*, *88*, Article 101592.

Shimada, N. (2022). An experimental study on strategic preference formation in two-sided matching markets. *ISER Discussion Papers No. 1169*, *1169*, 1–15.

Skiadas, C. (1997). Conditioning and aggregation of preferences. *Econometrica*, 347–367.

Sonnemans, J. (1998). Strategies of search. *Journal of Economic Behavior & Organization*, *35*(3), 309–332.

Staines, G. L., Tavris, C. A., & Jayaratne, T. E. (1974). The queen bee syndrome. *Psychology Today*, *7*(8), 55–60.

Strack, P., & Viefers, P. (2021). Too proud to stop: Regret in dynamic decisions. *Journal of the European Economic Association*, *19*(1), 165–199.

Sugden, R., Wang, M., & Zizzo, D. J. (2019). Take it or leave it: Experimental evidence on the effect of time-limited offers on consumer behaviour. *Journal of Economic Behavior & Organization*, *168*, 1–23.

Ulyssea, G. (2018). Firms, informality, and development: Theory and evidence from Brazil. *American Economic Review*, *108*(8), 2015–47.

UNDP. (2022). *Human Development Report 2021-22*. United Nations Development Programme.

Van Ginkel, W. P., & Van Knippenberg, D. (2008). Group information elaboration and group decision making: The role of shared task representations. *Organizational Behavior and Human Decision Processes*, *105*(1), 82–97.

Weidmann, B., & Deming, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, *89*(6), 2637–2657.

Younkin, P., & Kuppuswamy, V. (2018). The colorblind crowd? Founder race and performance in crowdfunding. *Management Science*, *64*(7), 3269–3287.

Zeelenberg, M. (1999). Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making*, *12*(2), 93.

Zhang, Y. (2023). Discrimination in the Venture Capital Industry: Evidence from Field Experiments. *Working Paper*.