Strategic interaction and social information

Essays in behavioural economics

Brütt, K.

Publication date
2023

Document Version
Final published version

This thesis contributes to the behavioural economics of strategic interactions and the study of how social information shapes these interactions in distinct settings. In evaluating the role of social information in economic interactions, behavioural economics allows for biases and limits in human information processing and considers the role of social information beyond its instrumental value, acknowledging that forces such as social preferences and social norms affect the impact of such information. Employing laboratory experiments, theoretical models, and quasi-experimental designs, this thesis considers how variations in the informational and strategic environment impact the selfishness of groups, experimentation in teams, and wage negotiations.

Katharina Brütt holds a BSc in Management, Philosophy & Economics from the Frankfurt School of Finance & Management (2016) and an MPhil degree in Economics from the Tinbergen Institute (2018). In 2018, she joined the Center for Research in Experimental Economics and Political Decision Making (CREED) at the University of Amsterdam as a PhD student under the supervision of Arthur Schram and Joep Sonnemans. Katharina currently works as an Assistant Professor at the Vrije Universiteit Amsterdam.

Strategic Interaction and Social Information: Essays in Behavioural Economics
Katharina Brütt

811

tinbergen institute
GRADUATE PROGRAM | Erasmus VU

Strategic Interaction and Social Information: Essays in Behavioural Economics

Katharina Brütt

Universiteit van Amsterdam

Strategic Interaction and Social Information: Essays in Behavioural Economics

ISBN: 978 90 361 070 20

Cover illustration: Katharina Brütt using Midjourney

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **811** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Strategic Interaction and Social Information:
Essays in Behavioural Economics

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties ingestelde commissie,
in het openbaar te verdedigen in de Aula der Universiteit
op vrijdag 17 maart 2023, te 14.00 uur

door Katharina Brütt
geboren te Hamburg

Promotiecommissie

Promotores:	prof. dr. A.J.H.C. Schram	Universiteit van Amsterdam
	prof. dr. J.H. Sonnemans	Universiteit van Amsterdam
Overige leden:	dr. H. Lee	Universiteit van Amsterdam
	prof. dr. T.J.S. Offerman	Universiteit van Amsterdam
	dr. J.J. van der Weele	Universiteit van Amsterdam
	prof. dr. H. Oosterbeek	Universiteit van Amsterdam
	prof. dr. M.C. Villeval	University of Lyon
	dr. C.E. Rott	Vrije Universiteit Amsterdam

Faculteit Economie en Bedrijfskunde

Acknowledgement of financial support

The research for this doctoral thesis received financial assistance from NWO Research Talent Grant 406.18.501, the Research Priority Area Behavioral Economics of the University of Amsterdam, the 'A Sustainable Future' platform of University of Amsterdam, and the Amsterdam Center for Behavioural Change of the University of Amsterdam.

ACKNOWLEDGEMENTS

I first arrived in Amsterdam more than six years ago. At that time, the defense of my PhD thesis seemed like an uncertain and rather unlikely possibility, so I am beyond happy that the day has come! This defense now closes one chapter of my life here, but I am excited to see what Amsterdam still has in store for me. There are many people whose support and encouragement made this challenging journey both joyful and a success. I want to thank you all.

First, I want to express my gratitude to my two advisors, Arthur and Joep. What started with a very brief meeting surprisingly quickly resulted in the two of you becoming my MPhil and PhD supervisors, a joint grant application, and a collaboration that I am very glad about. You both allowed me to work independently and to explore my own interests, guiding me in determining what makes a question interesting and how to best answer these questions. Arthur, I was always impressed by your detailed and thoughtful feedback and replies that seemed to appear within seconds. Joep, I have learned a lot from your experience in designing experiments and overcoming design challenges in clever ways. I also immensely enjoyed visiting Arthur at EUI in Florence, and Joep's open door always allowed me to quickly drop by to discuss any doubts. Thank you both! I would also like to thank all my committee members for their time in considering this dissertation and for attending my thesis defense.

My journey in Amsterdam started with two intense years at TI. I am grateful for the friends I have made who helped me so much throughout this period, with chats over coffee and shared anger over problem sets. All starting at TI, I am particularly happy to have met Davide, Thao, Wally, Adam, Rik, and Aslı. I am thrilled that so many of us keep the connection with Amsterdam, so I am very much looking forward to many more game nights and shared meals. Thank you Alina, Luísa, and Eva for your support throughout the years. I hope for still many more coffee dates to come!

To everyone at CREED, thank you for making the years of my PhD this great, and for the fun memories – whether it's exciting chats about research, playing foosball, or having drinks and dinner together. I am happy to remain so close to this fantastic community. In particular, I want to thank Theo, Jeroen, Giorgia, and Joël for their sharp feedback and their encouraging words throughout the years. Ayşe, thank you for the countless zoom meetings during the pandemic to share frustrations and for your excellent qualities as a food guide. Junze and Alejandro, thank you for being great office mates, shying away from no distraction to chat! A big shout-out also to all my other fellow PhD students at CREED throughout the years. Stephan, Andrej, Kostas, Johan, Chris, Ollie, and Linh, from Tuesday's trivia to many foosball games, CREED wouldn't have been the same without you. Thank you also to Margarita, Julia, Aljaž, Ivan, Ailko, Natalie, Dianna, Jan, Oda, and Maria for shared conversations and drinks on so many occasions.

Uri, thank you for giving me the opportunity to spend time in San Diego, being a great host, and for all the feedback you provided! I also want to thank the people I met in San Diego, especially Nathalie, Alex, Daniel, and René, for making the time in San Diego extra special.

Research is clearly more fun when done together with others. Therefore, I really want to thank my amazing collaborators, Huaiping, Chiara and Andrea, for the many hours spent together, in person or online. I have learned a lot from you! I hope for many fruitful and fun in-person interactions to come, Chiara and Andrea (when are we doing the retreat?!), and for mutual visits in Amsterdam and wherever you will end up, Huaiping.

The finishing touches of this thesis took place at the Vrije Universiteit Amsterdam. Thank you, Klarita, for making this possible and being so welcoming. I look forward to hopefully many successful collaborations in the future and am excited to be part of VU now.

I am also grateful for my friends in Germany, especially Mona and Laura, who always give me a nice glimpse into life outside of the bubble that is academia. Thank you for being part of why returning to Hamburg still feels like coming home.

A huge 'Thank you' to my family, in particular to my parents, who have always supported me in so many ways and never minded the long drive to Amsterdam (too much) to visit. Over the years of my PhD journey, you were always curious about what I was doing, showed full trust that I would succeed, but understood at the same time that when we are together, I don't mind sometimes leaving the talk about research behind.

Last but not least, my biggest 'Thank you' goes out to Andi. I could not have wished for a better partner – during my PhD and beyond. I am quite certain that I would have quit at some point throughout the years if it hadn't been for your constant reassurance and support. Thank you for always being there for me, making me laugh, and always

reminding me of what is most important in life. I am so excited for the future together with you and seeing what kinds of adventures it holds.

CONTENTS

CHAPTER

1

INTRODUCTION

This thesis is composed of three essays on the role of social information in strategic interactions. Employing laboratory experiments, theoretical models, and quasi-experimental designs, this thesis considers how variations in the informational and strategic environment impact the selfishness of groups, experimentation in teams, and wage negotiations.

Behavioral economics has, by now, a long-standing tradition of including realistic elements of human behavior in the analysis of strategic interactions (see, e.g., Crawford, 1997, for an early overview). While *strategic interactions* are clearly defined in economics as situations in which the payoff of one economic agent is dependent upon the choices of others (Black et al., 2009), no clear definition of *social information* has emerged. It is commonly understood as information about the previous actions of others (Coffman et al., 2017). I will follow a slightly broader conceptualization of social information, referring to information about the actions as well as the consequences and outcomes of others' actions.

In evaluating the role of social information in economic interactions, behavioural economics allows for biases and limits in human information processing, which are important when assessing how individuals react to the signals generated by others' actions. For instance, in social learning contexts, such as those studied by Anderson and Holt (1997), individuals tend to discount the predictions of others and put, compared to standard economic predictions, excessive weight on their own private signals (Weizsäcker, 2010). Furthermore, as its name suggests, social information is inherently 'social'. Behavioral economics takes this seriously, considering the role of social information beyond its instrumental value for decision making, and acknowledging that forces such as social preferences and concerns for adhering to social norms affect the impact of such information (Shang and Croson, 2009). These aspects are essential in strategic interactions, the focus of this thesis, where individuals can influence each others' payoffs.

The impact of social information is well established in the public goods literature. In laboratory experiments, providing social information can induce conditional cooperation (Fischbacher et al., 2001). Also outside of the laboratory, social information increases the provision of public goods, whether these public goods are charity donations (Shang and Croson, 2009), or contributions to movie rating websites (Chen et al., 2010). However, the impact of social information goes beyond choices in public goods settings. For example, social information influences labour market outcomes (Coffman et al., 2017) and energy consumption (Allcott and Rogers, 2014).

This forms the starting point of my thesis. Moving beyond traditional public goods settings, how does the information about others' actions and outcomes affect strategic interactions? This thesis will consider different aspects of the provision of social information. Using controlled laboratory experiments, I exogenously vary how easily this information is available, the strategic motives to generate information, and the usefulness of social information in distinct strategic environments.

Chapter 2 studies the diffusion of responsibility in endogenously-formed groups in a laboratory experiment. One of the core questions in behavioural economics is why and when individuals do not always act in their own economic self-interest, but instead are willing to give up their resources to help others. To answer this question, it is essential to understand which type of environments are conducive to such pro-social behaviour.

This chapter focuses on one important dimension of such an environment, namely whether decisions are made individually or by groups. Across different types of decision situations, from dictator games to prisoner dilemmas, the behavioural economic literature has documented that groups are often more selfish than individuals (Schopler et al., 1995; Luhan et al., 2009) and therefore behave more in line with standard economic predictions. Group decisions differ from individual decisions by the degree of responsibility individuals bear for the outcomes of their choices. If decisions are made in groups, responsibility is shared, leading to a so-called 'diffusion of responsibility'. This force generates more selfish behaviour (see, e.g., Behnk et al., 2022; Falk et al., 2020).

In this chapter, social information is ingrained in the default that determines the outcomes of group decisions if decisions are not made unanimously. The default provides information on the selfishness of the strategic environment that individuals may interact in. A selfish default informs individuals that the outcome of the group decision is selfish as soon as one group member votes in favour of this option, while a pro-social default communicates that the majority needs to vote in favour of the selfish option for this to be implemented.

This variation in defaults creates exogenous variation in the expected degree of responsibility diffusion for a selfish choice – a key dimension when individuals choose whether to decide as part of a group. With a pro-social default, every individual's vote in a group is pivotal in case a selfish choice is implemented, as everyone's individual vote could guarantee the pro-social outcome. The reverse is true with a selfish default. Here, responsibility is more easily diffused, since an individual selfish vote is not necessarily pivotal for the outcome of a selfish decision. First, we study how this variation in the expected degree of responsibility diffusion affects the extent of selfish behaviour. Second, the laboratory experiment examines how the impact of the possibility to diffuse responsibility differs by the process of group formation, contrasting settings where group membership is exogenous to settings where individuals can self-select into groups.

This chapter underlines the importance of group formation processes for the outcomes of group interactions. While the effect of the diffusion of responsibility on the extent of selfish choices vanishes with repetition in exogenously-formed groups, the effect is persistent and amplified in endogenously-formed groups. We document a striking selection pattern that causes this amplification. Pro-social individuals frequently opt out of group decision making if the default is selfish, but are more likely to enter groups if the default is pro-social. In contrast, selfish individuals are attracted by group decision making if the default is selfish. Thus, they seek an environment that allows for the diffusion of responsibility. The resulting group composition of endogenously-formed groups produces outcomes that show a larger degree of responsibility diffusion.

Chapter 3 continues to focus on interactions in groups, turning to how experimentation in teams can be encouraged. Some of society's most pressing issues, such as climate change or global pandemics, call for innovative new technologies (Bouckaert et al., 2021). Teams play an increasingly dominant role in producing these breakthrough technologies (Wuchty et al., 2007) and innovation is praised as a 'team sport' by leading management consultancies (Banholzer et al., 2019). This chapter explores what type of environments encourage team experimentation, that is, the process of trial and error at the core of innovation.

When experimenting in teams, social information is key: individuals learn from each others' experimentation efforts and need to consider how their experimentation will affect their team members' future actions. Therefore, this chapter also studies the supply of social information. Team members generate information by experimenting with a project of uncertain quality. The lack of positive news, which may arrive as a breakthrough, provides valuable information that informs later-stage experimentation efforts. If there is no breakthrough despite high experimentation efforts, future experimentation with this project is discouraged.

The strategic considerations of individuals who experiment in teams are shaped by two types of externalities (Bonatti and Hörner, 2011). First, in the type of teamwork I study, the expected payoffs of the team members are increasing in each team member's experimentation efforts, so there is a traditional free-riding problem in teams (Holmstrom, 1982). Second, when agents experiment with a project of uncertain quality, their experimentation effort generates feedback that all team members can capitalize on in the future. This creates an informational externality.

This chapter sets out to understand the determinants of strategic experimentation in teams in a laboratory experiment, varying two dimensions of the experimentation environment. First, I vary whether experimentation is joint or separate. If team members are experimenting with one joint project, one team member's experimentation provides an informational externality for the other team members. Since all team members work on one project, the information generated by each member is equally informative for all. In contrast, if experimentation involves separate projects, the projects' qualities are independent and there is no informational externality.

As a second dimension, I vary the observability of experimentation effort. Since only joint projects involve informational externalities, the observability of experimentation has distinct theoretical effects in these two settings. With joint experimentation, the observability of experimentation efforts discourages early experimentation, as high experimentation efforts without a breakthrough are a strong signal of working on a low-quality project. The goal to keep fellow team members optimistic creates an incentive to reduce early experimentation if it is observable. This is not the case when team members work on separate projects.

These predictions are in stark contrast to the experimental findings. Irrespective of whether experimentation is joint or separate, the team members experiment more if their experimentation is observable. Despite the informational externality that joint experimentation entails, joint experimentation also increases experimentation efforts compared to separate experimentation. The findings in this chapter are consistent with individuals leading by example to create norms of high experimentation. The observed experimentation levels are closer to the efficient benchmark than the equilibrium predictions. Therefore, it is key in strategic experimentation to consider not only the traditional informational value of social information. Next to the financial and informational incentives accounted for in standard economic models, experimentation also reacts to behavioral forces present in many other public goods settings.

Chapter 4 examines another policy-relevant application of the use of social information. In this chapter, I ask how social information, particularly information on others' wages and performance, can help decrease the gender pay gap. Wage transparency regulation has become an increasingly popular policy tool in response to the persistence of the gender pay gap, such as the 13% gap in the EU in 2020 (Eurostat, 2021). Ten EU member states have already adopted some form of pay transparency regulation. Despite its apparent popularity, there is little evidence on what determines the effectiveness of such regulation. Some legislations seem successful (see, e.g., Blundell, 2021; Bennedsen et al., 2022), while others are not (see, e.g., Gulyas et al., forthcoming; Cullen and Pakzad-Hurson, 2021).

By studying wage transparency, this chapter emphasizes the role of social information in correcting misguided beliefs. Employees often have incorrect beliefs about their co-workers wages. Importantly, women are, on average, more pessimistic about their co-workers' wages than men (Briel et al., 2022). This pattern is already present in the earnings expectations of adolescents (Boneva et al., 2022). At the same time, women are less confident in their performance (see, e.g., Niederle and Vesterlund, 2007). Information on the wages and performance of others may shift both types of beliefs, and, given the gender gap in these beliefs, do so more for women than men. Importantly, this type of social information also entails an element of social comparison – which may affect its uptake.

Our study investigates the recent introduction of a pay transparency regulation in Germany and complements this with a laboratory experiment to study the mechanisms that may limit its effectiveness. Upon request, firms with more than 200 employees need to provide information on the wages of employees performing comparable tasks. Using large-scale administrative data, we exploit exogenous variation induced by the threshold in firm size and variation across time to provide causal evidence on the effectiveness of this legislation. There is no evidence that this legislation affects wages or the gender wage gap.

In a laboratory experiment that mimics wage negotiations between a worker and a firm, exogenously, compared to endogenously, provided wage information does increase overall wages. Furthermore, information about a comparable worker's performance also increases workers' wages. However, there is no evidence that this effect is gender-specific. Furthermore, women enter negations less frequently if wage information is available, increasing the gender gap in negotiation entry. This suggests that poorly-designed transparency regulation may even backfire.

The results in this chapter provide important insight into the role of social information in negotiations. It is important not only whether such information is available per se, but how easily it is available and what type of other information is at hand at the same time. Zooming in on how social information can help correct misguided beliefs, we observe that if wage information is provided in isolation, individuals do not only update their beliefs about the wages of their comparable workers. Instead, they attribute surprisingly high wages to higher-than-expected performance. Therefore, it is crucial to consider how providing social information along one dimension can impact beliefs along other dimensions.

CHAPTER

2

ENDOGENOUS GROUP FORMATION AND RESPONSIBILITY DIFFUSION: AN EXPERIMENTAL STUDY.

This chapter is based on Brütt et al. (2020).

2.1 Introduction

This paper revisits the role of responsibility diffusion as an explanation of why groups make more selfish decisions than individuals. It also explores how endogenous group formation affects the degree to which group decision making induces selfish behaviour and whether group entry is exploited as a tool to diffuse responsibility.

The literature in behavioural economics has established that decisions by exogenously formed groups often differ from those made by individuals facing the same environment. For overviews, see Charness and Sutter (2012) or Kugler et al. (2012). In many circumstances, group decisions more closely follow standard economic theory and are more selfish compared to individual decisions. Examples include Schopler et al. (1995) for prisoner-dilemma games, Bornstein and Yaniv (1998) for ultimatum games and Luhan et al. (2009) in dictator games.[1] In search of an explanation for this phenomenon, the more recent literature examines the role of individual responsibility for immoral group actions. Individuals in a group collectively share the responsibility, which is thereby diffused. This is theoretically studied by for instance Rothenhäusler et al. (2018) and experimentally in Behnk et al. (2022) and Falk et al. (2020). If individuals do not expect to be pivotal decision makers, this allows every actor to believe that they are not fully responsible for the final outcome of the group decision-making process. This reduces incentives to act pro-socially or to act in accordance with a costly moral norm. In a related phenomenon, increasing the number of bystanders witnessing an emergency makes any single bystander less inclined to help, since individuals feel less responsible (Darley and Latané, 1968; Latané and Darley, 1968).

While there is a large strand of literature that compares the choices of individual decisions to group decisions in exogenously formed groups, little is known about how these choices would compare to an environment where the group is allowed to form endogenously. The literature thereby neglects the obvious fact that groups are evolving entities, which, in naturally occurring situations, often form voluntarily. From boards overseeing firms' activities and political committees determining a party's policy to families that jointly make household decisions, group decision making is ubiquitous, but often a choice in itself. Some people might simply prefer to make certain decisions in teams.

Our exploration starts with studying whether conclusions about selfish behaviour in (exogenously formed) groups carry over to an environment with endogenous group formation. Moreover, if group decision making facilitates selfish choices by reducing the disutility subjects experience due to the responsibility they bear, do individuals anticipate this? Does this make individuals prefer group decision making? Many studies emphasise the effects of so-called 'moral wiggle room', introduced initially by Dana

[1] Cason and Mui (1997), however, observe the opposite for dictator games, with groups acting less selfishly than individuals.

et al. (2007). This refers to the possibility of reducing individual moral behaviour by blurring the relationship between actions and consequences. Because group decision making may be seen as offering such wiggle room, we want to investigate whether individuals actively seize such opportunities by joining a group. If this is the case, a third question arises, about *who* seizes these opportunities. Consider a context that involves a trade-off between own and others' wellbeing. Can we expect a selection effect when there is voluntary group membership? That is, will 'selfish' individuals – those benefitting from a reduction in their responsibility – disproportionately join groups? This would imply that endogenously formed groups make (even) more selfish decisions than those formed exogenously. The study of exogenously formed groups would then, in fact, underestimate the true extent of selfish behaviour by groups.

To study how responsibility diffusion can serve as a determinant of group entry, a novel experimental design is introduced. This adopts a simple binary dictator game with a pro-social and a selfish option. The game is played by either individuals or groups of three players. We distinguish between groups that are exogenously formed and groups that are endogenously formed. Endogenously formed groups are composed of individuals stating a preference for group decision making over individual decision making.

Our design focuses on one central dimension of responsibility diffusion, which is a reduction of individual pivotality in a group decision. The idea is that one does not bear the full responsibility of a collective action if one's individual vote is not decisive in the implementation of that action. We will take the distance to being pivotal (Engl, 2018) as a measure of the extent of this decisiveness. In short, this means that an agent is deemed less responsible for a group choice if more votes need to change to make her vote pivotal in the implementation of a selfish outcome. This measure captures how as part of a group, one's vote may not be influential in the implementation of a certain decision. If one were to change one's vote, this would then not affect the outcome of the group decision-making process.

An example of this is voting in political elections. Every individual only bears a small fraction of responsibility for the outcome of an election and is usually far from being pivotal. The effect of responsibility diffusion, however, is likely to be asymmetric, which we are going to assume throughout this paper. Agents only want to diffuse their responsibility for the implementation of a selfish decision, not for a pro-social one. Given this asymmetric effect, we only consider the effect of being pivotal for the implementation of a selfish outcome.

We isolate this effect of being pivotal for the implementation of a selfish allocation by requiring unanimous decisions and varying the default option in case of non-unanimity. With a pro-social default, the group decision-making process excludes the possibility of diffusion of responsibility. The selfish outcome is only implemented if this is unanimously agreed upon, resulting in everyone's vote being pivotal and everyone bearing the full responsibility. In contrast, a selfish default facilitates diffusion of responsibility, since only one individual is required to vote in support of the selfish outcome for this to be realised. If at least one additional individual votes in favour of the selfish outcome, the agent is not pivotal for the realised outcome. This default then reduces the individual probability of being pivotal for this selfish outcome and therefore individual responsibility for this allocation.[2]

In this way, this design allows us to isolate the effects of changing pivotality for selfish outcomes on group decisions and to study how this effect is influenced by selection due to endogenous group entry.[3] Diffusion of responsibility likely has multiple dimensions. Note that we are interested in a dimension of responsibility diffusion that goes beyond a feeling of shared guilt. According to our understanding of shared guilt, an individual's disutility from causing a harmful outcome diminishes purely because this is done together with others. In this case, the decision makers share the burden of acting selfishly so the guilt is shared. Shared guilt is then unrelated to an individual's own impact on the group decision. Our design holds guilt sharing constant. In contrast, the dimension of responsibility diffusion we examine is directly linked to a person's individual impact on the decision. In the cases we study, the presence of other decision makers reduces this individual impact when there is a selfish default, but not when there is a pro-social default. Guilt is shared irrespective of the default.

To provide a benchmark for behaviour observed in the experiment, we employ a simple categorisation of individuals based on their degree of responsibility aversion. This predicts differences in selection and voting behaviour. From this, we derive hypotheses about the influence of responsibility aversion on selection behaviour and decisions in endogenously and exogenously formed groups. As we observe behaviour in repeated choices, our experiment can also speak to the effects of responsibility diffusion once subjects develop experience with their decision environment.

Our experimental data provide strong evidence for the importance of the group formation process. For exogenously formed groups, we find evidence that the diffusion of responsibility generates selfish choices in initial periods. This effect diminishes with repetition, which contrasts with arguments put forward in the previous literature. With endogenously formed groups, on the other hand, there are striking differences throughout the course of the experiment between environments where responsibility can be diffused and those where it cannot. If the selfish choice represents the default, more subjects vote selfishly than if the pro-social option serves as a default. Our design allows us to attribute this effect directly to the fact that group participation

[2] We abstract from social-image concerns here. In the discussion of Result 2, we will address why these concerns do not seem to be an important force here.

[3] Below, we discuss an additional selection effect. A pro-social default allows group members to force a pro-social outcome, irrespective of their group members' votes. This is not possible with a selfish default. Thus, a pro-social default may attract individuals that wish to promote pro-sociality.

is voluntary. To understand group decision making, both the possibility to reduce individual responsibility for selfish actions and a group's formation process should therefore be taken into account.

We continue and examine the role of self-selection of individuals into groups with specific defaults. First, we detect differences across types in the impact of a selfish default on selection decisions. More specifically, we confirm that types who vote selfishly if responsibility diffusion facilitates this, often seek responsibility diffusion in groups. In other words, a selfish group default has a positive influence on the propensity of selfish individuals to join a group.

Second, the experiment also provides evidence that individuals who – regardless of the type of decision making – choose the pro-social alternative, enter groups if they can guarantee that the pro-social outcome is implemented. This suggests a second motive driving group entry (aside from seeking responsibility diffusion). We find this effect to be stronger and more robust than the selection effect of selfish types. This finding is consistent with the social identity literature (e.g. Tajfel, 1974). If a pro-social default attracts group entry of pro-socials, this allows them to select into an environment with like-minded individuals. This identity effect seems to be more pronounced for pro-socials than for selfish individuals.

Finally, our study allows us to directly compare the decisions of individuals to group decisions in an environment where groups form voluntarily. As a result of the described self-selection, we find that differences between group and individual decisions particularly emerge in environments with a pro-social group default.

The remainder of this paper is organised as follows: First, we relate this study to the literature in section 2.2. Next, we explain the experimental design in section 2.3 and outline the hypotheses to be tested in section 2.4. Section 2.5 presents the results. Last, we will conclude and briefly discuss the findings in section 2.6.

2.2 Related literature

This study is at the intersection of three strands of literature in behavioural economics; the literature on responsibility diffusion and guilt sharing in groups, the literature on endogenous group formation, and the literature on sorting.

2.2.1 Responsibility diffusion and guilt sharing

Several authors have studied how responsibility can be assigned to single actors that are part of a collective decision-making process and how responsibility may be diffused in groups. Recent theoretical contributions define and formalise the concept of responsibility. Bartling and Fischbacher (2012) define a measure of responsibility that assigns to each agent a degree of responsibility for the implementation of an allocation that depends on the impact of the individual's action on its realisation probability. They do so in the context of decision delegation. More closely related to our framework is Engl (2018). He investigates how responsibility can be attributed if multiple agents are involved in a joint decision-making process. He proposes to take the distance to being pivotal as a measure of an individual agent's (ex-post) responsibility. This is defined as the number of changes in individuals' votes needed to generate a situation where this agent's choice can be decisive in changing the outcome. We will apply this measure below. Finally, Rothenhäusler et al. (2018) develop a model to identify the causal effects of guilt sharing on moral transgression with heterogeneous moral costs. Despite the similarities, shared guilt differs in their set-up to our notion of responsibility diffusion as it depends on the absolute number of supporters of immoral behaviour, independent of how decisive an individual's vote is.

Aside from these theoretical contributions, there is experimental work on responsibility in group decisions. Bartling et al. (2015) is somewhat related to our research. They study the impact of pivotality on responsibility attribution. The authors show in a sequential voting game that voters who are pivotal for the implementation of an unfair distribution of resources are punished more harshly than non-pivotal voters. Our experiment does not study punishment. Instead, our design allows us to directly link the degree of responsibility to the selfishness of a decision. Further experimental evidence is obtained in a labour market setting by Charness (2000, 2004). His results highlight that effort provision in a gift-exchange game is adversely affected if the responsibility for the outcome can be assigned to others or to an external process. Once again, this provides evidence for the importance of responsibility attribution.

Closest to our work is the experimental literature that studies the diffusion of responsibility in group decision making. One such study is Dana et al. (2007), who compare decisions of individuals and two-person groups (pairs) in a binary dictator game. A pair only implements the selfish outcome if both group members vote in favour of this option. As a consequence, both individuals in a pair are pivotal for an unfair outcome. Note that here is no variation in the dimension of responsibility diffusion we are interested in, which is linked to pivotality. While Dana et al. (2007) keep the pivotality in case of a selfish outcome constant, this is exactly what we want to vary and the effect we want to isolate. Yet, the authors observe that pairs are significantly more selfish than individuals.

Similarly, Behnk et al. (2022) observe more selfish decisions by pairs than by individuals in sender-receiver games. However, pairs can only send a deceptive message if both members agree to do so. This implies that every group member is pivotal for sending a deceptive message. As in Dana et al. (2007), the results suggest that the higher selfishness in pairs is the result of sharing guilt rather than reducing pivotality.

Selfish decisions by groups are also observed by Irlenbusch and Saxler (2019). They investigate the diffusion of responsibility by comparing decisions by individuals and groups to trade in a market environment where trade causes a negative externality to a third party. Other things equal, groups are more likely to engage in such trade. Given the veto power of an individual group member, an individual vote is again pivotal in groups.

All in all, group decisions tend to be less pro-social than individuals' in these studies. Nevertheless, this cannot be attributed to responsibility diffusion in the sense of reducing pivotality, but rather to shared guilt due to the mere presence of another selfish decider. In all of these studies, an individual's vote contributes just as much to the implementation of a selfish action in a group as it does in individual decision making. This will hold for any environment where unanimity is required for the group to choose selfishly. See Kocher et al. (2017) for an overview of alternative explanations for selfish group behaviour.

Falk et al. (2020) study the diffusion of responsibility in groups in an original manner. Participants in groups of eight face the choice between receiving money at the cost of voting to kill eight mice or foregoing the monetary payoff and not voting to kill the mice. A group implements the first option as soon as one group member votes to kill the mice. In contrast to the studies mentioned before, this means that group members are unlikely to be pivotal for the decision to kill the mice (such pivotality only occurs if all others forgo the money). The results show that the diffusion of responsibility in group choices results in significantly more mice being killed. An interesting observation in Falk et al. (2020) is that an individual's willingness to choose the selfish option decreases with the perceived likelihood of being pivotal. This supports the hypothesis that responsibility diffusion drives differences between individual and group decisions here.

All in all, this literature shows that groups make more selfish decisions than individuals and that a reduction of responsibility may be one of the factors that plays a role. We add to this literature in two important ways. First, we introduce a new mechanism that allows us to directly manipulate the possibility to reduce pivotality and isolate it from shared guilt. Second, endogenising group entry allows us to study whether agents exploit group decision making to make more selfish choices and whether this results in endogenously formed groups being even more selfish than suggested by the literature.

2.2.2 Endogenous group formation

To date, endogenous group formation has primarily been studied in public good experiments. For instance, Ahn et al. (2008) report a significant impact of the group formation process on public good provision. Both Ehrhart and Keser (1999) and Brekke et al. (2011) demonstrate that endogenously formed groups can sustain higher levels of public good provision because more pro-social individuals select into groups.

Aside from the literature that studies the endogenous formation of institutions in exogenously formed groups, such as Sutter et al. (2010), there is an emerging literature that looks at the endogenous formation of groups with punishment institutions. Kosfeld et al. (2009) show that this endogenous formation increases public good provision. Gürerk et al. (2006, 2014) and Fehr and Williams (2018) show that self-selection into groups and endogenous migration across groups with distinct punishment institutions fosters cooperation. Nicklisch et al. (2016) add that the observability of contribution levels in a public good game encourages individuals to join groups with punishment opportunities. Robbett (2014) further stresses that next to the ability to enter groups with different institutions, the ability to then shape these institutions is crucial.

Similarly, in weakest link games Riedl et al. (2015) and Chen (2017) find individuals choosing the set of players and the group to interact with is effective in promoting efficient coordination. In contests, Herbst et al. (2015) find that self-selection into groups has a significant impact on effort, both through intensified free-riding and increased in-group favouritism. The field study by Hamilton et al. (2003) considers self-selection into group production, finding (somewhat surprisingly) that there is no adverse selection of team members.

Importantly, groups are not decision makers in any of these studies. Instead, individuals make decisions that affect others in their group. Nevertheless, such findings suggest that the neglect of the endogeneity of group membership is a serious gap in the existing literature. The only study we are aware of that considers endogenous group formation in the context of group decision making is Kocher et al. (2006), who do so for beauty contests. They find that while endogenously formed groups perform better than individual decision makers, high ability subjects are more likely to opt for individual decision making. The context we propose here involves an entirely different strategic environment. Furthermore, to the best of our knowledge, ours is the first study that directly compares distributive *decisions by* endogenously and exogenously formed groups, as opposed to *behaviour within* those groups.

2.2.3 Sorting

Few studies consider the effect of sorting or self-selection on the decision environment. One notable exception is Dana et al. (2006). They demonstrate that a substantial fraction of subjects opt out of playing a dictator game when this ensures that the receiver stays uninformed about the option of playing this game. Also Lazear et al. (2012) show that some types of individuals sort into environments that do not provide the opportunity to share their earnings. This self-selection significantly alters the degree of pro-social behaviour in a dictator game.

Grossman and Van der Weele (2017) investigate sorting of dictators into different information states concerning the payoffs of receivers. They find substantially more pro-social behaviour of dictators who self-selected into environments with knowledge of the receiver's payoff.

Similar to our experiment, these studies underline that sorting provides a mechanism that can decrease the extent of pro-social behaviour. We add to this by considering sorting into various types of groups as such a mechanism.

2.3 Experimental design

Our experimental design aims to shed light on the effects of responsibility diffusion and endogenous group formation on differences in group and individual choices. In particular, it will allow us to study selection effects in response to the possibility to reduce pivotality for selfish decisions.

The design employs a binary dictator game with the following two options:

$$A : (10, 0)$$
$$B : (6, 6)$$

where the first number depicts the dictator's and the second number the receiver's payoff in experimental tokens. *A* represents the selfish option and *B* the pro-social option. This dictator choice involves no strategic interaction. As a consequence, choices are independent of beliefs about other players' behaviour. Throughout the experiment, the dictator and receiver roles are neutrally labelled as Player 1 and Player 2, respectively.

2.3.1 Treatments

The experiment uses a within-subject design. There are five distinct treatments, varying along two dimensions. First, treatments differ with respect to how decisions are made and second, they differ in the way groups are formed, see Table 2.1. Groups always consist of three individuals.

Exogenous group formation	Endogenous group formation
• Individual decision making (*ExoInd*) • Group decision making *A* default (*ExoGroupA*) • Group decision making *B* default (*ExoGroupB*)	• *A* default (*EndoA*) • *B* default (*EndoB*)

Table 2.1: Treatment overview

All subjects make decisions in the dictator role (how payoffs are determined is explained below). Decisions are made individually or in a group. To start, we explain the treatments in which groups are formed exogenously (as is the allocation to the individual decision-making treatment). Each participant is matched with one other individual, the receiver.

If decisions are made individually, as in *ExoInd*, subjects decide about their payoff and the payoff of the other participant they are matched with. If decisions are made in a group, as in *ExoGroupA* and *ExoGroupB*, a group jointly decides about their payoff and the payoff of another group, consisting of the three individuals the group members are matched with. Within matching groups of nine individuals, subjects are rematched into different groups in each period that involves group decision making. Subjects know that they are rematched in every group decision-making period, but are unaware of the matching groups. Each group member in the dictator and the receiver group obtains the payoffs corresponding to the implemented option. This ensures that group decisions do not differ from individual decisions in terms of per-capita incentives, which could alter the perception of the selfishness of choosing either option. The role of outcome-based explanations for different choices is therefore limited.

Group decisions are made by each group member simultaneously casting a vote. Every period, every individual has only one chance to cast a vote. A decision is implemented if group members unanimously vote in favour of it. There is no communication between group members. This is in line with other studies like Dana et al. (2007), Falk et al. (2020) and Behnk et al. (2022). If group members do not unanimously vote for either option, a default option is implemented. In *ExoGroupA* this default option is A, in *ExoGroupB* it is B.

This specification of the voting process serves as a mechanism to isolate the effects of the pivotality dimension of responsibility diffusion on group choices. If the pro-social outcome B is the default, then there is no possibility to diffuse responsibility, as in having a reduced probability of being pivotal for a selfish decision (compared to when the decision is made individually). As all group members need to agree on A for this to be implemented, every vote is pivotal if A is implemented and every group member bears the responsibility for A being the aggregate decision. In contrast, if A is the default, this allows for diffusion of responsibility. A vote in favour of A is here only pivotal for implementing A if the two other group members vote in favour of B. As soon as at least one other group member votes in favour of A, an individual vote in favour of A is no longer pivotal. In expectation, if we assume a positive probability of any individual voting in favour of A, more votes need to change if the default is A than if the default is B to change the outcome from selfish to pro-social. With default A, an agent voting in favour of A can expect to be further away from being pivotal for the selfish group choice A. This reduces the expected responsibility for implementing the selfish outcome. If one individual would change her vote, the selfish option *A* would still be implemented.

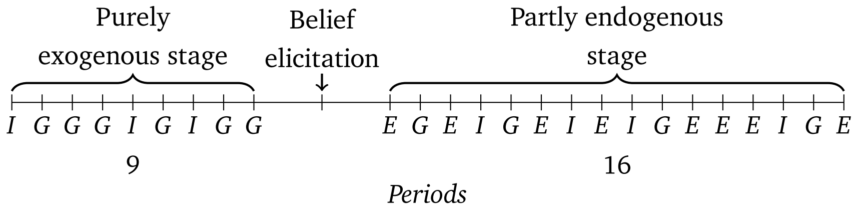
In contrast to Behnk et al. (2022) and Falk et al. (2020), analysing the effect of diffusion of responsibility here involves the comparison of group decisions in distinct environments. Both Behnk et al. (2022) and Falk et al. (2020) compare group to individual decisions, while we leave all aspects of group decision making constant except the default option.[4] This is especially important as the evidence in favour of a shift to more selfish behaviour in dictator games is inconclusive. Luhan et al. (2009) find more selfish behaviour by groups but Cason and Mui (1997) find the opposite. We apply the dictator game because its lack of strategic interaction provides a clean measure of pro-sociality. We prefer to compare groups to groups (as opposed to groups to individuals) because this allows us to directly capture variations in the diffusion of pivotality without confounds from other drivers of differences between group and individual decisions.

The two treatments in which groups are formed endogenously, *EndoA* and *EndoB*, involve decisions in two stages. First, individuals choose between group and individual decision making. When choosing, they know the default option for a group they would join. This is *A* in *EndoA* and *B* in *EndoB*. In the second stage, individuals face, depending on their first stage decision, one or two choices. If individual decision making is chosen, participants face the same choice as in *ExoInd*. If, instead, group decision making is preferred, subjects have to indicate, both, how they would vote if they indeed join a group and how they would decide if they cannot join a group for decision making. We use the strategy method here because this allows us to study groups that are formed purely voluntarily. In case the number of subjects that indicates a preference for group decision making is not a multiple of three, randomly selected participants who prefer group decision making are assigned to make the individual decision.[5] This is common knowledge. Out of nine subjects in a matching group, at most two subjects may have to decide individually instead of in a group, despite preferring group decision making.[6]

[4]As explained above, only *ExoGroupA* allows for a diffusion of responsibility; *ExoGroupB* provides agents with veto power against the implementation of selfish choices. If an agent favours a pro-social allocation of resources, a vote in favour of *B* guarantees this allocation in *ExoGroupB*.

[5]The design is not symmetric here. Subjects opting for individual decision making only face one question, because individual decision making is guaranteed if this is preferred. So their vote in case of joining a group would not have been incentivised. Since subjects may especially dislike inconsistencies between their individual and group decision in a simultaneous choice, our results provide a lower bound of the selection effect.

[6]Note that there is no need to exclude these data. Since we use the strategy method, subjects decide between *A* and *B* for both the case they end up as an individual decision maker and the case they end up as part of a group. They do so before they know whether they are an individual decision maker or part of a group. Both of these choices are properly incentivised, as it is possible to end up in either case.



I: Period with exogenously-determined individual decision making
G: Period with exogenously-determined group decision making
E: Period with endogenously-determined group or individual decision making

Figure 2.1: Timing

2.3.2 Belief elicitation

If subjects join groups because this allows them to diffuse responsibility, then the likelihood that a selfish individual will join a group is higher, the more other selfish choices she expects. To measure this, we elicit participants' beliefs about the number of selfish choices by others.[7] We do so for four distinct cases, distinguished by the default and whether the decision is individual or in a group.

Each participant is asked to indicate how many out of ten randomly sampled other subjects are expected to choose *A*. For each elicitation, subjects are endowed with one euro. The elicitation is incentivised by subtracting ten cents multiplied by the absolute difference between the estimated and observed frequencies from this endowment.[8] Asking for frequencies in this way instead of subjective probabilities yields a more accurate elicitation of beliefs (Schlag et al., 2015).

2.3.3 Timing

This experiment consists of 25 periods plus the belief elicitation stage. These 25 periods include multiple periods of each treatment. As illustrated in Figure 2.1, the experiment is split into an initial nine periods without endogenous group formation and an additional 16 periods that alternate in a fixed order between exogenous individual and group decision-making periods and endogenous group formation. The belief elicitation takes place in between these two sets of periods.[9]

The first part serves to allow subjects to gain experience both with individual and with group decision making in environments where *A* and *B* are the defaults. This part consists of three rounds of *ExoInd*, three rounds of *ExoGroupA* and three rounds of *ExoGroupB*. Comparing exogneous group decisions in the first part to those decisions in the later part of the experiment allows us to investigate whether the effect of

[7] Eliciting such beliefs also enables us to estimate whether participants believe that others use groups to diffuse responsibility.

[8] This linear scoring rule is used to facilitate understanding of the mechanism.

[9] Since the payoff of the belief elicitation only depends on *other* subjects' behaviour, this belief elicitation is not expected to affect behaviour in the second part of the experiment.

responsibility diffusion is stable across time.

The second part of the experiment comprises four periods of *EndoA* and *EndoB* each, two periods of *ExoGroupA* and *ExoGroupB* each and four periods of *ExoInd*. To reduce as much as possible any effects from the order of decisions taken, periods are organised in blocks of four in the second part. Each block is made up of one *EndoA* period, one *EndoB* period, one *ExoInd* period and one exogenous group decision-making period. The latter alternates between *ExoGroupA* and *ExoGroupB*. Our main analysis will focus on this second part of the experiment, since the level of experience is here similar in the between-treatment comparison.

At the end of the experiment, participants are randomly split into dictators and receivers. Only the decisions by those allocated to the dictator role are implemented. As a consequence, every subject's choices potentially determine someone else's payoff in each of the 25 rounds.

2.3.4 Procedures

The experiment is computerised using PHP (the programme is available upon request). It was run at the CREED laboratory of the University of Amsterdam. We ran 12 sessions in June and October 2018 with a total of 216 students.[10] Subjects with a variety of backgrounds were recruited. A questionnaire administered at the end of the experiment shows that 63% have a background in either economics or business and 50% are female. Each session consisted of 18 participants, constituting two matching groups. Subjects did not know the identity of the participants who were in their matching group. Upon arrival, subjects were randomly allocated to computer stations and further communication was prohibited. The instructions and images of the decisions screens are available in appendix 2.A and appendix 2.B, respectively. Understanding of the instruction was tested using practice questions (cf. appendix 2.A). Subjects were only allowed to move on and start the first period of the experiment if they answered all questions correctly.

Final payoffs consisted of a show-up fee of seven euros plus the earnings of the 25 periods and the belief elicitation. Each experimental point corresponds to a payoff of 0.08 euros. Average earnings were 21.71 euros; sessions lasted on average approximately 45 minutes.

[10] There was a pilot session before these 12. After the pilot, we introduced a small change in waiting times in endogenous periods. This guaranteed that individuals have no incentive to decide individually in order to avoid waiting for other participants.

2.4 Behavioural hypotheses

To derive testable hypotheses, we assume that, next to their monetary payoff, individuals care about the extent to which they are responsible for the implementation of a selfish outcome. The associated moral costs depend on two factors.

We call the first 'responsibility aversion'. This reflects an individual-specific parameter that captures the degree to which an individual is affected by being responsible for decreasing someone else's earnings in comparison to the alternative choice. In the experiment, individuals with distinct responsibility aversion would each attach a different weight to diminishing the receiver's earnings from six to zero by choosing A instead of B. The higher the responsibility aversion, the higher are the moral costs for diminishing someone else's payoff.

The second factor reflects the 'degree of responsibility' for implementing a selfish outcome. More precisely, it captures the extent to which the individual concerned is pivotal in the decision to choose A. Holding the responsibility aversion constant, the disutility from decreasing someone else's payoff increases in the degree of responsibility. As introduced by Engl (2018), we take the distance to being pivotal as a measure of responsibility, with a higher distance reflecting lower responsibility.[11] An agent's distance to being pivotal is defined as the number of votes, including that agent's vote, that need to change such that her vote would be pivotal for changing the outcome. The expected distance to being pivotal is then the expected number of votes that need to change. It gives the simple binary concept of pivotality as the determinant of individual responsibility in collective action more nuance.

As argued above, this distance to being pivotal for choosing A is expected to be higher in group decisions with default A than with default B and than in individual decisions. If, for example, A is the default and everyone votes in favour of A, the distance to being pivotal is three. Everyone needs to change their vote for the outcome to change. If B is the default and everyone votes in favour of A, the distance to being pivotal is one. So even if everyone votes selfishly, the distance to being pivotal remains low with B as the default. If B is the default, the expected distance to being pivotal for implementing A is therefore independent of agents' beliefs about others' behaviour.

To derive testable hypotheses, we classify individuals based on their degree of responsibility aversion into three types. A given level of responsibility aversion leads to specific choice patterns in groups with varying defaults. Realise that decisions of individual decision makers should be the same as their votes in groups with B as the default, since in both cases they are fully pivotal in case of a selfish decision and responsibility thus cannot be diffused.[12] In Appendix 2.C we present and analyse a

[11]While we use the measure of Engl (2018), it is not our purpose to test his model. Such a test would require a different design.

[12]We have no way to know whether subjects realise that they are as pivotal for implementing A if

decision-theoretic model that formalises the classification presented here. This model also provides a formal underpinning for the hypotheses presented below.

Individuals with a low level of responsibility aversion choose A irrespective of the group's default; even if responsibility cannot be diffused, they prefer to vote for the selfish outcome. Note that these individuals are not necessarily completely indifferent to the consequences of their actions for others' payoffs. However, their responsibility aversion is not high enough to make them forego the monetary gains associated with the selfish option under either default. Because of their choice pattern, we call these individuals 'selfish'. In periods with endogenous group entry and A as the default, these individuals will choose to decide in the group because they can assure a selfish group choice but still might get some benefit out of the diffusion of responsibility. The expected distance to being pivotal for implementing A is increased, if there is a positive probability of one other individual voting A, while A can still be guaranteed. If B is the default, they cannot guarantee their preferred outcome A and by lack of responsibility diffusion, there is no advantage to joining a group. They therefore prefer to decide individually and then choose A.

At the other extreme are those who have a high level of responsibility aversion. This leads them to choose B even with default A, so they choose B irrespective of the default. With A as the default, the expected distance to being pivotal for allocation A is higher, but this does not suffice to compensate for the moral costs that result from being (even only partly) responsible for a selfish outcome. Because of this voting pattern, we call them 'pro-social'. These individuals prefer to decide individually when A is the default because the consequences of being partly responsible for a selfish group choice (A) weighs too heavily on them. Therefore, they prefer to decide individually and implement the pro-social B. When B is the default, they are indifferent about joining a group, because they can always ensure outcome B either way.

Finally, there are individuals with an intermediate level of responsibility aversion. Given their degree of responsibility aversion, they choose A when A is the default and B when B is the default. Their responsibility aversion is high enough to kick in and make them choose B when B is the default and the distance to being pivotal for implementing A is not increased; the moral costs of voting selfishly are too high if responsibility is not diffused. Yet responsibility aversion is low enough to make them choose A if the expected distance to being pivotal for the selfish choice is higher with A as the default. Responsibility is then diffused sufficiently such that the higher payoff offsets the moral costs of being partly responsible for someone else's low payoff. We call this type of individuals a 'switcher'. Such an individual will be indifferent about joining a group when B is the default because her choices will ensure the outcome B

this is done in a group with B as the default as they are individually. However, our practice questions do ensure that subjects know that they only override the default if everyone votes in favour of doing so.

Responsibility Aversion	Group decision A default	Group decision B default	Type	Selection A default	Selection B default
low	A	A	selfish	group	individual
intermediate	A	B	switcher	group or individual	indifferent
high	B	B	pro-social	individual	indifferent

Notes: The level of responsibility aversion (column 1) determines decisions in groups with default A (column 2) and default B (column 3) which results in the type classification (column 4). Classification determines selection of environment (columns 5 and 6).

Table 2.2: Classification of different types of individuals

in any case. When A is the default, she may choose to decide individually or to join a group, depending on the precise value of her responsibility aversion. If responsibility aversion is relatively low, diffusion of responsibility still makes it sufficiently attractive to enter groups and exploit this feature. Higher levels of responsibility aversion ensure that voting individually for B dominates being selfish in a group, as the costs of bearing the responsibility for selfish outcomes are too high despite responsibility diffusion.

Table 2.2 summarises this discussion.

2.4.1 Experimental hypotheses

Our classification yields the following testable hypotheses.

Hypothesis 1. *In exogenously formed groups, option A will be chosen more often if A is the default than if B is the default.*

This assumes a positive mass in the type distribution on the switching type. In that case, this hypothesis follows directly from Table 2.2 by observing that switchers choose A (B) when A (B) is the default, while other types do not change their behaviour based on the default. Because A is the default which allows for diffusion of responsibility, this hypothesis predicts that the diffusion of responsibility generates more selfish decisions. This is because it decreases the moral costs of making a selfish choice. In our experiment, we expect the share of A choices to be higher in treatment *ExoGroupA* than in *ExoGroupB*. This first hypothesis serves as a robustness check of previous findings of Behnk et al. (2022) and Falk et al. (2020) using a different identification mechanism for the effects of diffusion of responsibility.

Hypothesis 2. *The difference in the share of votes in favour of A between groups with A as the default and groups with B as the default is larger in endogenously than in exogenously formed groups.*

Given the theoretical selection patterns depicted in Table 2.2, individuals with lower responsibility aversion – that is, those more likely to choose the selfish option –

select into groups when this entails the possibility to reduce responsibility for selfish actions. In addition, those with high responsibility aversion (who are likely to choose the pro-social option *B*) select out of groups when responsibility can be diffused. Together, this means that endogenously formed groups with default *A* have a higher fraction of selfish individuals than the randomly formed exogenous groups with this default. Similarly, a group with default *B* will not be joined by selfish individuals, so the fraction of selfish types in these groups will be lower than in exogenous groups. This selection of selfish types into endogenously formed groups with a selfish default and out of endogenously formed groups with a pro-social default yields the conclusion that endogeneity amplifies the differences between the defaults. Hence, comparing the difference in selfish decisions in treatments *EndoA* and *EndoB* to this difference in *ExoGroupA* and *ExoGroupB*, we expect the former to be higher. This assumes a positive mass on at least two types with different selection choices under default *A*.

Hypothesis 3. *Individuals classified as selfish join groups more often if A is the default.*

This follows directly from Table 2.2 and the discussion following Hypothesis 2. The fraction of selfish group entrants in *EndoA* is thus predicted to exceed the fraction of selfish group entrants in *EndoB*.

Hypothesis 4. *Individuals classified as pro-social join groups more often if B is the default.*

We predict that the opposite of Hypothesis 3 holds true for types that do not take advantage of the diffusion of responsibility, the pro-social types. This is, however, more of a conjecture than a formal result. Though pro-social types prefer individual decision making if *A* is the default (cf. Table 2.2), they are indifferent when *B* is the default. The hypothesis follows if we assume that indifference yields a positive probability of joining a group. In that case, we predict that the fraction of pro-social types that join groups is higher in *EndoB* than in *EndoA*.

Note that the behaviour predicted in Table 2.2 is deterministic in the sense that someone either joins a group or does not. Adding noise to decisions would mean that the choices become probabilistic. That is, as the level of responsibility aversion increases, the likelihood to join a group with default *A* (*B*) decreases (increases). Then, the ideas underlying hypotheses 3 and 4 can also be tested using the elicited beliefs. In the reasoning underlying Table 2.2, a selfish subject's or switcher's propensity to join groups in *EndoA* will increase in her beliefs about the share of selfish votes in groups because this increases the likelihood that she will be able to diffuse responsibility. The same holds for switchers. The reverse holds for pro-social types, whose propensity to join groups should decrease in their beliefs about the share of selfish votes because more of such votes will increase the likelihood of a selfish outcome.

2.5 Experimental results

As mentioned above, we have data from 12 sessions (for a total of 216 participants) in a within-subject design. We start by outlining the methodology underlying the data analysis.

2.5.1 Methodology

We use both parametric and non-parametric techniques. Because of the dependence of observations within matching groups, observations are averaged at the matching group level whenever using non-parametric tests. This results in 24 observations per treatment. We apply Fisher-Pitman permutation tests for paired replicates (FPP) when making within-subject comparisons. When comparing choices across different types of individuals, we use Fisher-Pitman tests for independent samples (FPI) (Kaiser, 2007). These permutation tests are non-parametric and do not require distributional assumptions (Siegel and Castellan, 1981). At the same time, permutation tests are more powerful than traditional non-parametric techniques such as the Wilcoxon signed-rank and Mann-Whitney U tests. Moir (1998) demonstrates in a Monte Carlo analysis that permutation tests allow for more robust inference with small samples, as discussed in Schram et al. (2019).

In the parametric analysis, standard errors are clustered at the matching-group level to ensure consistent estimates despite dependent observations. p-values reported in the main text stem from these clustered standard errors. To account for a possible downward bias in the estimated residuals due to the small number of clusters (Cameron and Miller, 2015), we include a robustness check using score wild cluster bootstrap to calculate p-values based on Kline and Santos (2012) in all tables.

We categorise subjects into types along the lines of Table 2.2. To do so, we use choices in periods with exogenous group decisions in the second part of the experiment (constituting two choices each with defaults A and B). If subjects vote for A, irrespective of the default, they are categorised as selfish, and if they always vote for B, they are the pro-social type. If out of the four responses considered, subjects more often vote for A if A is the default than if B is the default, they are assigned to be a switcher. This classification captures 78% of the subjects, with 17% of all subjects being categorised as selfish, 24% as switcher and 37% as pro-social.

To investigate whether our analysis requires a trend correction, Figure 2.2 shows the share of individuals deciding for or voting for A over the 25 periods. This fraction is close to 50% in all periods, which allows testing the outlined hypotheses without fearing boundary effects. The mild downward trend in the fraction of selfish choices or votes (Spearman's $\rho = -0.047$; $p < 0.001$) suggests that experience slightly matters. Unless stated otherwise, we therefore consider only decisions in the second part of the

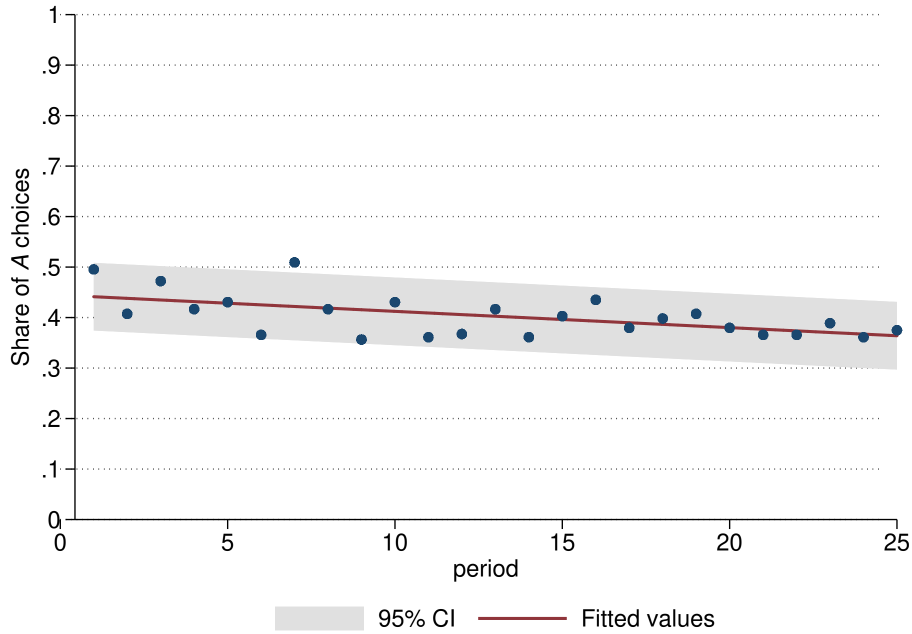


Figure 2.2: Share of selfish choices per period

experiment, where there is no detectable downward trend (Spearman's $\rho = -0.017$; $p = 0.324$)[13]. This is especially important in the comparison of endogenous and exogenous periods to ensure that subjects in exogenous and endogenous rounds have a comparable level of experience. Here, treatments are spread out evenly.

2.5.2 Results

Table 2.3 presents summary statistics for the main variables of interest (the fraction of A choices and the choice to enter a group).

In the upper panels, we compare the fraction of A choices and the choice to enter a group across defaults. In the lower panels, the table displays the results of a difference-in-difference analysis. The inequalities indicated are differences in the fraction of A choices or differences in the choice to enter a group across defaults. In the lower-left panel, the cross-default difference in the fraction of A choices is compared between endogenously and exogenously formed groups. In the lower-right right panel, differences in the choice to enter a group are compared between individuals classified or not classified ("others") as a certain type.[14]

Result 1. *In exogenously formed groups, option A is not chosen significantly more often when A is the default than when B is the default.*

[13] For the two correlation tests observations are not averaged at the matching group level.

[14] For example, the selfish join the group 15 percentage points more often when A is the default than when B is the default. Together, those not classified as selfish join a group 8 percentage points less often when A is the default than when B is the default. The diff-in-diff between +15 percentage points and -8 percentage points is statistically significant with $p < 0.01$.

	A votes			Group Entry				
	Exo		Endo	Total		Selfish	Switching	Pro-social
Overall	0.37	$>^{*}$	0.31	0.40		0.28	0.39	0.47
A default	0.38	$>$	0.38	0.38		0.35	0.40	0.39
	$\vee$	$\wedge^{***}$	$\vee^{***}$	$\wedge^{*}$		$\vee^{**}$	$\vee$	$\wedge^{***}$
B default	0.36	$>^{**}$	0.25	0.42		0.20	0.38	0.56
				Others				
Difference				-0.08	$<^{***}$	0.15		
between	0.02	$<^{***}$	0.13	-0.06	$<^{**}$		0.02	
defaults				0.03	$>^{***}$			-0.17
Observations	863		692	1727		288	416	640

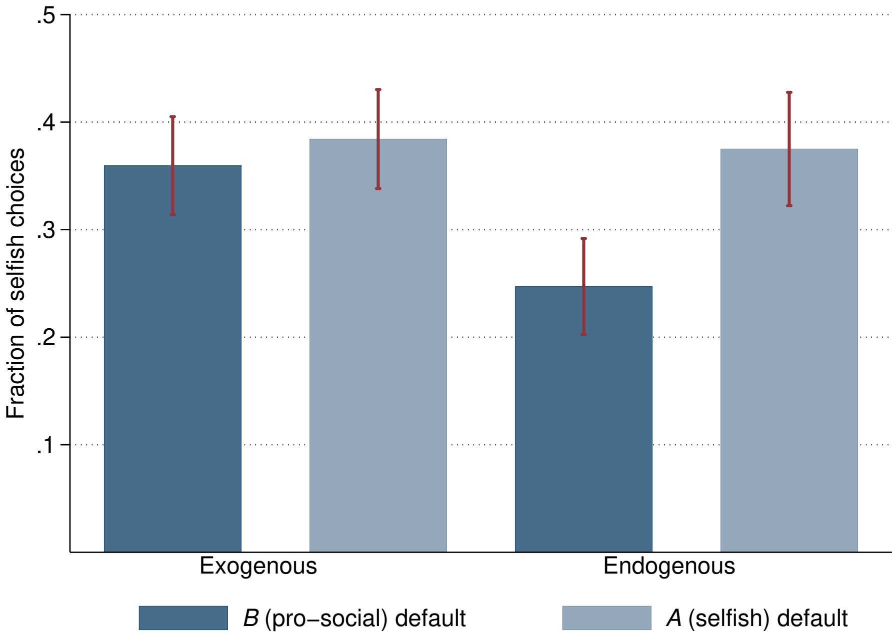
Notes: Cells report the average fraction of *A* votes in groups and the average fraction of individuals choosing to enter a group. Differences of means and differences in differences are tested with Fisher-Pitman permutation tests using matching group averages. The numbers in the lower panels are explained in the main text. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.3: Summary statistics and tests on selfish votes and group entry

Hypothesis 1 can thus not be confirmed when focusing on decisions in the second part of the experiment (see below, however, for results on the first part). Using the numbers from Table 2.3, Figure 2.3 illustrates the fraction of votes in favour of *A* across the group treatments. Although the fraction of votes in favour of *A* is slightly higher in exogenously formed groups when *A* is the default (0.38) than when *B* is the default (0.36), this difference is not statistically significant (FPP; $p = 0.373$).

This result is in contrast to the findings of Behnk et al. (2022) and Falk et al. (2020). Recall that these studies measure responsibility diffusion by comparing individual's decisions when alone to their decisions when in a group. Dimensions of responsibility diffusion other than a change in pivotality might, however, cause distinct behaviour in the two settings. Of course, we cannot exclude the possibility that for some reason responsibility diffusion plays a role in the settings studied by Behnk et al. (2022) and Falk et al. (2020), while it does not in our experimental setting. Nevertheless, we can conclude that changes in an individual's pivotality do not always trigger more selfish behaviour in groups.

Given that our design comprises 25 periods, we can investigate whether and how responsibility diffusion matters at different stages of the experiment. To investigate whether our treatment variation is successful in varying the subjects' perception of responsibility, we explore behaviour of subjects in exogenously formed groups in the first part of the experiment. This allows us to see whether the increase in the probability of being pivotal initially leads to more selfish behaviour by those subjects susceptible to a variation in responsibility. In contrast to the results in the second part of the experiment, we observe that in the first part subjects vote significantly more selfishly in groups with a selfish default. Over the whole course of the experiment, 41% of the



Notes: Fraction of votes in favour of A across treatments in part 2. Error bars indicate 95% confidence interval based on individual observations.

Figure 2.3: Comparison of the fraction of selfish choices across treatments

votes are selfish in exogenously formed groups with a selfish default, which is significantly more than the 37% in such groups with a pro-social default (FPP; $p = 0.043$). This effect is even stronger if we only consider periods in the first part of the experiment (44% vs. 38%; FPP; $p = 0.015$). Table 2.4 summarises this.

The effect of a reduction in the probability of being pivotal thus seems to diminish with repetition. This means that our results are consistent with the previous literature, such as Falk et al. (2020) and Dana et al. (2007). These previous studies apply one-shot interactions or only a few repetitions. Falk et al. (2020) find in their 'charity paradigm' that subjects are more selfish in a second period, where subjects are able to form more accurate beliefs about pivotality. Similarly, it is conceivable that with more repetitions subjects start to realise that they are using the diffusion of responsibility to justify their behaviour.

Given that we observe a difference between group decisions with different defaults in the early periods of the experiment, we conclude that our design is successful in varying subjects' perceived diffusion of responsibility. Yet, this variation does not result in a significant change in behaviour in later periods, where subjects start to learn more about their decision environment.[15]

[15]To check whether the lack of differences in the exogenous treatments may be due to subjects' strive for consistency, we examine the intra-individual correlations in exogenous periods in the second part of the experiment. These are only moderate (Pearson's $\rho = 0.4$ between group decision periods with A and B as the default; Pearson's $\rho = 0.42$ between group decision periods with A as the default and individual decisions; Pearson's $\rho = 0.49$ between group decision periods with B as the default and

	A votes					
	Only 1st part			1st and 2nd part		
	Exo Group		Exo Ind	Exo Group		Exo Ind
A default	0.44	$<$		0.41	$<$	
	$\vee^{**}$		0.48	$\vee^{**}$		0.44
B default	0.38	$<^{***}$		0.37	$<^{***}$	
Observations	648		648	1080		1513

Notes: Cells show per default the average fraction of *A* votes in exogenously formed groups and exogenously imposed individual decision making. Differences in means are tested with Fisher-Pitman permutation tests using matching group averages. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2.4: Comparison of votes including first part of the experiment

Strikingly, highly significant differences in voting behaviour across defaults do exist in the second part in endogenously formed groups (FPP; $p < 0.001$). As illustrated in Figure 2.3, the effect size of introducing a selfish default is much larger in endogenous compared to exogenous groups. Note that differences in exogenously formed groups can only be attributed to a default-induced change in the pivotality for selfish outcomes. We have concluded that this change in pivotality plays no significant role for the exogenous case in the second part. Differences in endogenously formed groups, on the other hand, may stem from either this change in pivotality or from a differential selection that alters the group composition. Because pivotality on its own does not matter here (as seen in the exogenous periods), these differences must stem from selection effects, possibly in response to changes in pivotality. We discuss these selection effects below.

The difference between endogenously and exogenously formed groups gives the second result.

Result 2. *The difference in the share of votes in favour of A between groups with A as the default and groups with B as the default is larger in endogenously than in exogenously formed groups.*

This result confirms Hypothesis 2. The cross-default difference in the number of selfish votes is clearly larger in endogenously formed groups than in exogenously formed groups (0.13 versus 0.02, see Table 2.3). The difference-in-difference is highly significant (FPP; $p = 0.003$). This means that having a choice between individual and group decision making results in a more pronounced role of defaults. This provides clear evidence that group formation processes matter and should be taken into account when studying group decisions.

We next use the results in Table 2.3 to consider the selection mechanisms driving these amplified differences. Note first that individuals in endogenously formed

individual decisions).

groups with a selfish default, A, are not significantly more selfish than those in exogenously formed groups with this default (both choose A approximately 38% of the time; FPP; $p = 0.538$). Second, individuals in endogenously formed groups with a pro-social default, B, vote for B significantly more often than individuals in exogenously formed groups with this default (75% versus 64%; FPP; $p = 0.017$, see Table 2.3). Interestingly, this suggests that the amplification of differences in endogenously formed groups is not primarily driven by the possibility to diffuse responsibility for selfish choices, but by the possibility to ensure non-selfish allocations. This is in line with the expected selection of pro-social types depicted in Table 2.2, that predicts that pro-social types do to not join groups with A as the default, yielding relatively fewer selfish choices if B is the default.

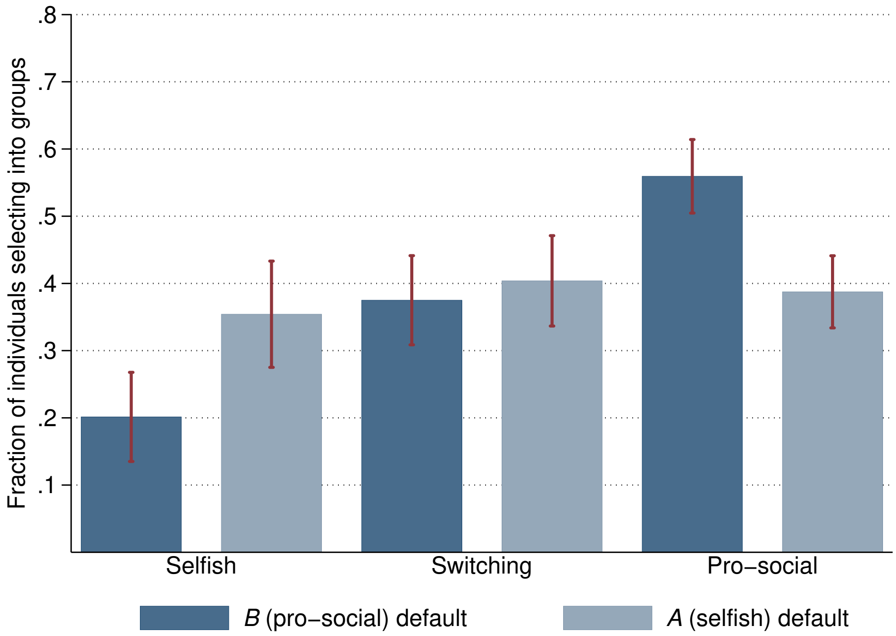
This result also demonstrates the limited role of self-image concerns in this setting. Individual group entry decisions are not primarily motivated by the opportunity to caste a pro-social vote at no actual cost in groups with A as the default. This would have implied more pro-social votes in endogenously formed groups with default A, which is contrary to what we find.

We will now investigate the type-specific selection behaviour in more detail. Overall, individuals join groups 40% of the time, documenting a willingness to forego autonomy for the sake of group decision making. At the aggregate level, we observe a marginally significant effect of the default on selection into group decision making (FPP; $p = 0.097$), with default B being more attractive. This result, however, hides type-specific heterogeneity. Figure 2.4 clearly shows that types exhibit differences in their group entry response to a change in the default. These differences are discussed in the next results.

Result 3. *Individuals classified as selfish join groups significantly more often if A is the default.*

This confirms Hypothesis 3. Individuals categorised as selfish join groups in 35% of the cases when A is the default (allowing for the diffusion of responsibility) and in only 20% of the cases with default B, see Figure 2.4. The difference is significant (FPP; $p = 0.015$). We see that those individuals that benefit from a selfish default enter groups more often if their selfish vote's pivotality can be reduced. A selfish default, therefore, mainly attracts individuals who, given their choice pattern, actually benefit from the moral wiggle room it provides.

Interestingly, as group entry is voluntary, group members are responsible for being part of a group that allows them to diffuse responsibility. In this way, joining a group for the opportunity to act selfishly could be seen as an immoral act in itself. In principle, this could diminish the moral wiggle room that appears once one is in the group. Our result implies, however, that for selfish types the attraction of reducing their selfish vote's pivotality is strong enough to outweigh the fact that group mem-



Notes: Fraction of individuals selecting into groups with A or B as the default for different types. Error bars indicate 95% confidence intervals based on individual observations.

Figure 2.4: Selection into groups across types

bers deliberately selected into this environment. Therefore, the fact that voluntary group entry entails responsibility for those that enter the group does not prevent selfish individuals from exploiting this wiggle room. Their benefit from being less likely the pivotal selfish decision maker is large enough.

As Figure 2.4 shows, switchers do not respond to the default in their selection choice. We do not find any effect of the default on group entry; the difference between 40% joining with default A and 38% with default B is insignificant (FPP; $p = 0.204$). The final formal result describes pro-socials' selection choices.

Result 4. *Individuals classified as pro-social join groups significantly more often if B is the default.*

This result confirms Hypothesis 4. Figure 2.4 reveals that, in contrast to other types, pro-social types' frequency of joining a group is significantly lower when A is the group's default than when B is the default (39% versus 56%; FPP; $p = 0.002$). As predicted, this type of individual does not seem to want to free ride on others' selfish voting behaviour to benefit from a high payoff without bearing the responsibility for it. In fact, they join groups if, given the default, an equal split of resources is likely. This suggests that pro-social types join groups when they can guarantee an equitable split for a whole group of participants (they can ensure that the group decision is equitable if B is the default). Note from the bottom row of Table 2.3 that there are more pro-social types than selfish or switchers. With default B, the pro-socials also have a much higher probability of joining the group. As a consequence, a majority of the group members with default B is pro-social (56% of those we can classify). It is possible that these subjects also join groups to be among like-minded individuals. Individuals may derive a higher utility from acting pro-socially (choosing B) if this is done in a group they can identify with.[16] While selfish individuals might also prefer to be in a group with similar other individuals, this urge is likely stronger for pro-socials.

Aside from a comparison of means, we also see that the distributions of selection frequencies differ across types. Figure 2.5 shows the distribution of individual frequencies of selecting into groups.[17] The figure illustrates that the distribution of group entry frequencies differs significantly between selfish and pro-social types if the default is the pro-social B (Kolmogorov-Smirnov test; $p < 0.001$), but the distributions are much more similar if groups face a selfish default (Kolmogorov-Smirnov test; $p = 0.988$).

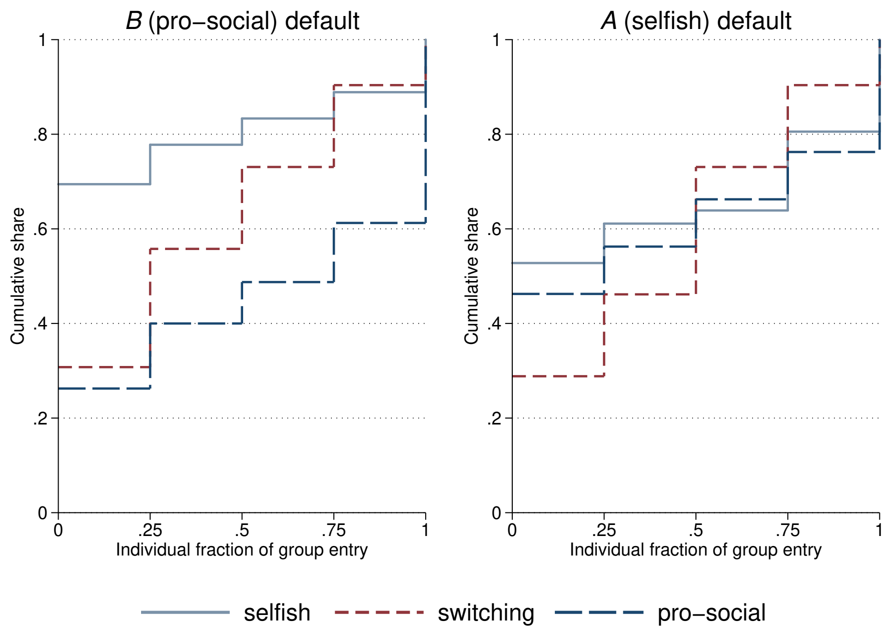
Self-selection therefore significantly alters the group composition in an environment with pro-social default, but not so much if the default is selfish.

To analyse how the effect of a default change on a type's group entry decision compares to other types' responses, we use the cross-default difference in the fractions of individuals selecting into groups as a measure to be compared across types. These fractions are given in the lower-right panel of Table 2.3. For instance, we compare the difference in the fraction of pro-social individuals that choose group decision making across defaults (−0.17) to this difference for individuals that are not pro-social (0.03). As shown in Table 2.3, pro-social types, compared to other participants, are affected significantly more by a default change. The selfish default A makes them select into groups much less than default B does and this difference is significantly different than the difference for the average other individual, who enters slightly more often with A (FPI; $p < 0.001$). In turn, both selfish types and switchers exhibit a higher positive effect of a selfish default on selecting into group decision making than others (FPI; $p < 0.001$ and $p = 0.029$, respectively). These numbers underline the heterogeneous effects that a change in default has across types.

For an estimate of the magnitude of the type-specific selection effects, Table 2.5 reports the logit regression results of group entry on the default and the respective

[16]As noted by an anonymous reviewer, joining a group and subsequently acting pro-socially could also be explained by outcome-based social preferences. Choosing B, subjects can guarantee that six subjects receive a payoff of six each, while A would only give three subjects a payoff of 10 each. Sufficiently strong Fehr and Schmidt (1999) or Charness and Rabin (2002) type of preferences would allow for this. However, as we observe pre-dominantly pro-social voting in these groups, the pro-social default B is implemented in 98% of the cases. An individual's vote is pivotal in only 10% of these cases. Joining a group, then, typically does not yield a fair outcome for more individuals. This is why we suggest a preference for joining a group of similar individuals as an explanation (on top of the desire to force a pro-social outcome). This would be in line with Hett et al. (2020), who show that individuals are willing to forego substantial monetary gains to be part of a group they identify with.

[17]Recall that there are four group entry choices per default. This means that entry can be selected 0%, 25%, 50%, 75% or 100% of the time.



Notes: Cumulative distribution of individual frequencies of joining groups by type and default.

Figure 2.5: Comparison of selection distribution across defaults

average marginal effects. For the selfish, the marginal effect of a selfish default illustrates that selfish types join groups on average 15.3% more often if the default is selfish ($p = 0.006$). In contrast, pro-social individuals join groups significantly less frequently, the average marginal effect of a pro-social default B is 17.2% ($p < 0.001$).

To study the impact of these selection patterns, we compare the fraction of selfish, switching and pro-social types in groups with A as the default to groups with B as the default, see Table 2.6. This highlights how the observed differences in selection behaviour translate into differences in the group composition. If we allow subjects to reduce the probability of being pivotal for a selfish outcome, the group composition is affected in two ways: First, there are more selfish individuals in groups with A as the default (16% vs. 8%; FPP; $p = 0.016$). Second, there are fewer pro-social individuals in groups with A as the default (38% vs. 49%; FPP; $p < 0.001$).

These results are all robust to using traditional non-parametric tests, the Mann-Whitney U test for comparisons of independent samples and the Wilcoxon signed-rank test for within-subject comparisons, as reported in Appendix 2.D. We also investigate type-specific behaviour under alternative classification mechanisms, with similar results, which are discussed in Appendix 2.E. In particular, the selection behaviour of pro-social types proves very robust to how individuals are classified.

Sample	(1) Total	(2) Selfish	(3) Switching	(4) Pro-social
	Entry	Entry	Entry	Entry
A default	-0.176*	0.777**	0.121	-0.696***
	(0.102)	(0.305)	(0.192)	(0.154)
	[0.0901]	[0.0170]	[0.5335]	[0.0010]
	*-0.042**	*0.153****	*0.029*	*-0.172****
Constant	-0.315***	-1.378***	-0.511***	0.239
	(0.120)	(0.302)	(0.191)	(0.203)
Observations	1727	288	416	640
Clusters	24	17	23	21

Notes: Logit regression estimating the effect of A being the default on group entry for different types. Standard errors clustered at the matching group level are in parentheses. In square brackets, score wild cluster bootstrap p-values adjusted for small samples are reported. Average marginal effects are reported in italics. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: The effect of A being the default on group entry

	fraction of type in group		
	Selfish	Switching	Pro-social
A default	0.16	0.26	0.38
	$\vee^{**}$	$\vee$	$\wedge^{***}$
B default	0.08	0.21	0.49
Groups	24	24	24

Notes: Cells report the fraction of group members that have a certain type in groups with A and B as the default. The unit of observation is the average per matching group. Differences are tested with Fisher-Pitman permutation tests using matching group averages. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.6: Composition of endogenously formed groups

2.5.3 Differences in individual and group behaviour

So far, we have exploited the specific characteristics of our design by considering individual behaviour in groups and their entry decision, both under distinct defaults. We can also directly compare individual decisions in isolation and in groups, as is common in much of the previous literature. We add to this by including endogenously formed groups in the comparison. Our results are summarised in Table 2.7.

The results show that the fraction of selfish choices is lowest (25%) when the default is B and groups are formed endogenously and highest (48%) for individuals who had the option to join such groups and chose not to. The difference is highly significant (FPI; $p = 0.001$). Both can be explained as follows. Recall that we previously observed frequent group entry by pro-socials in this environment, where they could impose their preference on the group. As a consequence, the groups have relatively many pro-socials and there are relatively many selfish amongst the remaining individual decision makers.

In contrast, voting in self-selected groups and by individual decision-makers is very similar if the default is the selfish choice A (38% and 39%, respectively; see Table 2.7).

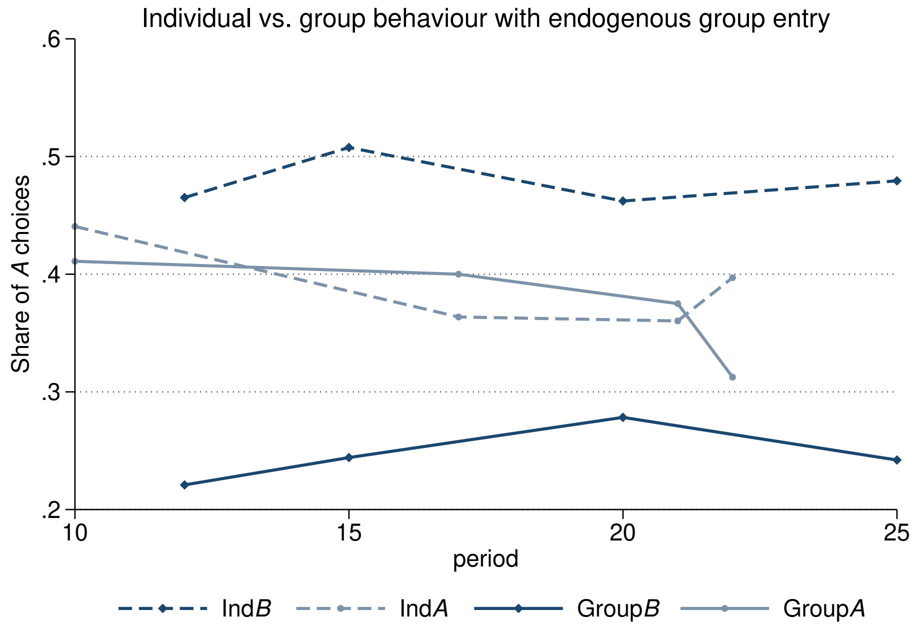
	A votes					
	Endo Group		Endo Ind	Exo Group		Exo Ind
A default	0.38	$<$	0.39	0.38	$<$	
	$\vee^{***}$	$\wedge^{***}$	$\wedge^{***}$	$\vee$		0.41
B default	0.25	$<^{***}$	0.48	0.36	$<$	
Observations	692		1035	863		865

Notes: Cells show per default the average fraction of *A* votes in endogenously formed groups, endogenously chosen individual decisions, exogenously formed groups, and exogenously imposed individual decision making. Differences in means are tested with Fisher-Pitman permutation tests using matching group averages. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2.7: Comparison of group and individual decisions

The difference is statistically insignificant (FPI; $p = 0.921$). This confirms our earlier conclusion that group entry based on a desire to diffuse responsibility is much weaker than entry to be with other pro-socials. To see whether these aggregate results hide any differences across rounds, Figure 2.6 shows the fraction of selfish choices for the four rounds in which a specific environment was implemented. This shows that the mean fraction of selfish choices is relatively constant across the four cases in each of the four situations an individual can end up in.

Contrary to previous findings in the existing literature, we cannot detect significant differences between individual and group behaviour with either default if groups are formed exogenously (FPP; $p = 0.127$ if B is the default, $p = 0.337$ if A is the default). Thus, as with the case of groups with different defaults, a reduction in the probability of being pivotal does not affect behaviour here. Table 2.4 even shows more selfish behaviour by individuals than by groups with a pro-social default if we take into account the first part of the experiment. This finding is in line with the comparison of group and individual behaviour in dictator games by Cason and Mui (1997) but in contrast to findings by Luhan et al. (2009) and Dana et al. (2007). 44% of votes in the exogenous individual decision-making treatment are selfish. This is statistically indistinguishable from the 41% in exogenously formed groups with A as the default, but higher than the 37% in such groups with B as the default (FPP; $p = 0.265$ and $p = 0.004$, respectively). The pattern we observe does not conform with our analysis in section 2.4. The expected distance to being pivotal is unchanged in groups with a pro-social default compared to individual decision making, while it is increased in groups with a selfish default. Recall, however, that we concluded that the distance to pivotality plays no role in the second part of the experiment, which we are considering here. Apparently, the direct comparison of group and individual decisions does not isolate the diffusion of responsibility. Other aspects, such as shared guilt, play a role in the comparison between individual and group choices. This reaffirms our argument in favour of comparing distinct types of group decisions that systematically vary different aspects of responsibility.



Notes: Lines connect the fractions of selfish votes for individuals deciding alone or in groups for varying defaults in periods with endogenous group formation.

Figure 2.6: Group vs. individual decisions depending on default

Although our results for the dictator game do not contrast sharply with the literature, they are noticeably different from results for experiments that allow for strategic interaction. One potential explanation is that we do not employ any means of communication. Studies like Schopler et al. (1995), Bornstein and Yaniv (1998) and Luhan et al. (2009), which find that groups are more selfish than individuals, allow for communication. It is possible that more direct contact with other group members through communication creates a stronger feeling of group identity (see e.g. Chen and Li, 2009), which could facilitate the use of moral wiggle room such as offered by shared guilt. This should not matter in our main analysis when comparing different types of group decisions. Nevertheless, it could be relevant in the comparison of individual and group decisions.

Moreover, the social psychology literature stresses that group decisions often amplify observed individual tendencies, the so-called group polarisation phenomenon (see e.g. Moscovici and Zavalloni, 1969). The fraction of votes in favour of the pro-social option *B* in the exogenous individual decision-making treatment is 59%. Thus, a majority votes in favour of the pro-social option. Arguably, this makes the pro-social option the socially desirable choice. Similar to Cason and Mui (1997), social comparison might lead individuals in a group to present themselves more favourably than the average subject. Here, this would then imply choosing pro-socially. This process then provides an explanation for the group polarisation phenomenon.

Finally, Figure 2.6 shows a clear impact of endogenous group formation on the comparison of individual and group decisions. Differences between individual and group behaviour are amplified when group entry is voluntary for groups with a pro-social default B (FPP; $p = 0.015$), but not with a selfish default A (FPP; $p = 0.524$). This once again points to subjects using group entry to be part of a group with equally pro-social individuals.

2.5.4 Beliefs

Voting behaviour and group entry patterns are largely in line with the hypotheses. Analysing beliefs provides further insights into the mechanisms through which the diffusion of responsibility affects these choices.

Elicited beliefs show that subjects believe that if groups are formed endogenously, individuals will behave in a manner consistent with responsibility diffusion. With default B, subjects believe that 39% will vote selfishly in endogenously formed groups, while they believe that 52% will do so if A is the default. The difference is statistically significant (FPP; $p < 0.001$).

As a further way to investigate whether people enter groups because they believe that this will allow them to diffuse responsibility, we correlate the fraction of times that a subject entered a group with default A (out of four opportunities) with their beliefs about others' choices. If more individuals are expected to vote selfishly, this increases the expected distance to being pivotal of someone voting selfishly herself.

We observe neither in aggregate nor for any of the studied types a significant correlation between beliefs and group entry with default A. Subjects' group entry decisions do not depend on the expected degree of responsibility diffusion. This is confirmed by regression analysis (see appendix 2.F). Again it appears that the ability to ensure pro-social outcomes is a more important criterion for group entry than responsibility diffusion.

Indeed, we observe that with a pro-social default B subjects expect more pro-social votes in groups (61%) than pro-social choices amongst individual decision makers (48%) (FPP; $p < 0.001$). In other words, those entering a group correctly anticipate a pro-social outcome. As a consequence, pro-social individuals do not appear to join groups to change the outcome of the group decision (which is expected to be pro-social anyway). Instead, they appear to enjoy joining a group that makes a fair decision. This effect can only be established because we consider endogenous group formation.

2.6 Conclusion

We demonstrate that endogenous group formation matters in the context of social behaviour. It amplifies the differences we observe between group decision making with and without the possibility to diffuse responsibility. In our study, the degree to which a group member's vote is pivotal matters if individuals self-select into group decision making. Since this is one of the first studies to investigate the effects of endogenous group entry on group decisions, our findings suggest a promising avenue for further research.

Moreover, the analysis shows that these amplified differences in endogenous groups are the result of type-specific selection behaviour. Interestingly, an environment that allows for diffusion of responsibility does not, in the aggregate, make group decisions more attractive. Especially pro-social individuals appear to dislike this kind of environment and exhibit a preference for an environment that allows them to join like-minded pro-social individuals in ensuring an equitable outcome. In contrast, selfish individuals are more prone to select into group decision making in an environment where diffusion of responsibility is possible. This implies that from a policy perspective, the desirability of giving individuals the option to team up before deciding depends not only on whether group decision making would allow for diffusion of responsibility, but also on the type of individuals involved. In an environment where pro-social choices are the default – for example, because they constitute the status-quo or the norm – pro-socials will want to join groups. In our experiments, this is seemingly driven by a preference for acting pro-socially in a group of like-minded individuals. Furthermore, our results imply that, for instance, a committee that decides as a team in an environment where diffusion of responsibility is possible is likely to repel pro-social types. Instead, this environment is likely to attract more selfish types. When creating the rules that govern decision making, such selection effects should be considered.

Furthermore, we add to the existing literature on responsibility diffusion by investigating the repetitive nature of choices where responsibility can be diffused. We identify that a reduction in the probability of being pivotal changes behaviour in exogenously formed groups mainly in early rounds. The effect diminishes with repetition.

We have also contrasted the behaviour of self-selected individual and group decision makers. An environment that allows individuals choosing to join a group to force the pro-social outcome on the group exhibits stark differences in the degree of selfish behaviour between these groups and individual decision makers. Again, this can be attributed to pro-social types being attracted by the group environment while selfish types are deterred from entering such groups. As a consequence, voluntary group entry both affects group and individual decision making.

APPENDIX

2.A Experimental instructions

2.A.1 First part

The experiment consists of two parts and in total of 25 periods. Part 2 will be explained after we have finished Part 1.

The instructions are simple, and if you follow them carefully, you might earn a considerable amount of money. Your earnings can depend on your decisions and may depend on other participants' decisions. You will be paid in private and in cash at the end of the experiment.

Your decisions in the experiment are private to you. We ask you to not communicate with other people during the experiment. Please refrain from verbally reacting to events that occur during the experiment. The use of mobile phones is not allowed during this experiment. If you have any questions, or need assistance of any kind, please raise your hand and an experimenter will come to you. If you talk, laugh, exclaim out loud, etc., you will be asked to leave and you will not be paid. We expect and appreciate your following of these rules.

Your task

Part 1 of the experiment will consist of **9 periods**.

The decision situation

This experiment concerns the implementation of distributions of experimental points. Two parties can benefit from the distributions of experimental points. One of the two parties will be called Player 1, the other party will be called the Player 2. Player 1 and Player 2 have different roles. In each period, Player 1 will decide about the distribution of experimental points. Player 2 does not make any decisions.

In this experiment, there are two possible allocations:

- A: Player 1 receives 10 points, Player 2 receives 0 points
- A: Player 1 receives 6 points, Player 2 receives 6 points

The two alternatives are summarised in the table below:

		Payoff	
		Player 1	Player 2
Choice	A	10	0
	B	6	6

During the experiment, you will both be in the role of Player 1 and of Player 2. Everyone will only be making decisions as Player 1, since Player 2 is passive.

After the experiment, it will be randomly determined which half of the participants are assigned the role of Player 1 and which half are assigned the role of Player 2. This determines which half gets their payoff from being Player 1 and which half gets their payoff from being Player 2. Thus, with equal probability, either all of your decisions are relevant and determine your payoffs and the payoffs of participants that are assigned to be Player 2's, or your payoff entirely depends on other participants' choices. Since you do not make any decisions as Player 2, your only actions throughout the experiment will be actions as Player 1.

Be aware that if you are chosen to receive payoffs from being Player 1 at the end of the experiment, your actions in each period determine the sole payoff of another participant. The person you are matched with would not receive any payoffs from their choices as Player 1.

Types of decision-making

In the first part of the experiment, as Player 1 you will face two types of decision-making:

1. Individual decision-making
2. Group decision-making

In every period, you will need to choose between an option A and an option B. At the beginning of each period, you will be told whether you will be making this decision in a group or as an individual. In the first 9 periods, you will make 3 decisions as an individual and 6 decisions in a group.

If you are making decisions as an individual, you decide whether you want option A or option B to be implemented. You will be coupled with another individual, Player 2. If you choose option A, you will receive 10 points and Player 2 will receive 0 points. If you choose B, you will both receive 6 points.

In periods in which you are making decisions in a group, you are randomly matched with two other participants and you jointly decide whether you want to implement option A or option B. If the decision is A, each member of your group will receive 10 points. If the decision is B, each member of your group will receive 6 points. In case of group decision-making, your group is coupled with another group of three (Player 2) who will all be affected by your group's decision. If your group's decision is A, each member of the other group will receive 0 points, if option B is chosen each member will get 6 points. Note that the difference between individual and group decisions is that the group decision is implemented for everyone in your own group and everyone in the group with whom you are coupled. The following overview summarises the consequences of the decisions.

	Individual decision-making	**Group decision-making**
Option A	You receive 10 points. Player 2 (an individual) receives 0 points.	Every member of your group receives 10 points. Every member of Player 2 (a group) receives 0 points.
Option B	You receive 6 points. Player 2 (an individual) receives 6 points.	Every member of your group receives 6 points. Every member of Player 2 (a group) receives 6 points.

The group decision-making procedure is as follows: Each of the three group members will individually cast a vote. A decision will be implemented if the vote is unanimous, meaning that all three group members voted for the same option. In case the vote is not unanimous, a default option will be implemented. The default option is either A or B. You will know the default option before you cast your vote.

You will either be assigned the role of deciding between option A and B (Player 1) or the role of the individual or group to which Player 1 is coupled (Player 2).

If you are Player 1, then each time you are making a decision in a group, the group composition will change. This means that it is unlikely that your fellow group members are exactly the same any two times you make a group decision. If you are Player 2, the decisions of your group members will also not matter and therefore under no circumstances affect your earnings. Your group members are also assigned the role of Player 2 if you are Player 2. Therefore, you will also never affect your group member's earnings in case they are Player 2.

Throughout the experiment, no one will ever learn the identity of their group members or the identity of either the Player 1 or the Player 2 they are paired with.

Example

In this example, imagine a participant is assigned to make decisions in a group. The default is A. The participant is matched with two other individuals, the group members of the participant. Together, they fulfil the role of Player 1. The two group members both vote for B, while the participant votes for A. Hence, the vote is not unanimous: Not all group members voted for the same option. The default option, A, is therefore implemented.

The group with the role of Player 1 will receive 10 points each in this period, the group with the role of Player 2 they are matched with will receive 0 points in this period.

Payment

We will not tell you whether you are Player 1 or Player 2 in today's experiment. Everyone will make choices as if they are Player 1, which may well be the case. We have randomly divided you into two halves. After the final period, the computer will randomly determine which half will be assigned to be Player 1 and which half will be Player 2. Note that the randomisation is in no way related to any decisions made in this session. The decisions by the Player 1's will determine everybody's earnings today.

The experimental points you earned in each period of the experiment, either as Player 1 or Player 2, will be summed up and will be exchanged for money. The exchange rate is as follows: For each point you earn, you will receive € 0.08. Additionally, you will receive a show-up fee of € 7.

2.A.2 Practice question first part

We will now ask you some questions to check your understanding. You can always browse back to previous screens. When you have a question for the experimenter, please raise your hand.

1. If you are chosen to be Player 1 and you are deciding individually, how many points do you earn in a period in which you choose option A?

2. If you are chosen to be Player 1 and you are deciding individually, how many points does your Player 2 earn in a period if you choose option B?

3. If you are chosen to be Player 1 and you are making decisions in a group where B is the default, how many points do you earn individually in a period if one of your group members votes for A, one of your group members votes for B and you vote for A?

4. If you are chosen to be Player 1 and you are making decisions in a group where B is the default, how many points does each individual in the group of Player 2 who you are paired with earn in a period in which both of your group members vote for A and you vote for A?

2.A.3 Second part

Your task

Part 2 of the experiment will consist of **16 periods** and is similar to part 1 of the experiment.

Again, in each period, Player 1 will decide about the distribution of experimental points, that will at the end of the experiment be exchanged for money. Player 2 does not make any decisions. As before, in this experiment, there are two possible allocations:

- A: Player 1 receives 10 points, Player 2 receives 0 points
- A: Player 1 receives 6 points, Player 2 receives 6 points

The two alternatives are summarised in the table below:

		Payoff	
		Player 1	Player 2
Choice	A	10	0
	B	6	6

There are three different types of periods in this part:

1. Individual decision-making
2. Group decision-making
3. **Choice between individual and group decision-making**

At the beginning of each period, you are shown what type of period you are in. In total, you will face 4 periods of individual decision-making (type 1), 4 periods of group decision-making (type 2) and 8 periods in which you choose between individual and group decision-making (type 3). The two first types of periods are identical to what you have encountered in part 1 of the experiment, with the same rules applying here to individual and group decision-making as in part 1 of the experiment. Type 3 is new.

In periods in which you have the choice between deciding as an individual or group (type 3), you will first be asked whether you prefer to be part of a group or decide alone. In case you prefer to decide alone, you will be making your choice individually.

Groups will only be formed by those that prefer to decide in a group. In case you state that you prefer to be part of a group, you are likely to make your decision in a group. It is possible, however, that you will nevertheless have to decide as an individual; if the number of participants that wants to decide in a group is not a multiple of three, some individuals preferring the group need to choose individually. It is still in your best interest to state your preferred type of decision-making truthfully, since this will always increase the probability with which you will be assigned to your preferred type.

Before you choose between individual and group decision-making, you will always be told the default in a group that you can join. If you enter a group, your group will be paired with three individuals who are Player 2 in this period. If you opt for individual decision-making, you will be paired with one individual who is Player 2 in this period.

After choosing the type of decision-making, you face the choice of Player 1, similar to part 1. If you opted for individual decision-making, you have to decide between option A and B. If you prefer making decisions in a group, you will be asked to make both the decision on what you vote for in case you will be in group and on what you decide in case you are forced to decide individually after all.

Please note that after each period in which you choose between individual and group decision-making (type 3), you will wait for all your fellow participants before you go to the next period. Your waiting time is therefore independent of whether you choose individual or group decision-making.

Payment

As explained in part 1, you will not know until the end of the experiment whether you are Player 1 or Player 2 in this experiment. The randomisation at the end of the experiment will determine your role for both parts. This means that you are either Player 1 in both parts or Player 2 in both parts. The randomisation is in no way related to any decisions made in this session.

Recall that the experimental points you earned in each period of the experiment, either as Player 1 or as Player 2, will be summed up and exchanged for money. The exchange rate is as follows: For each point you earn, you will receive € 0.08. Additionally, you receive a show-up fee of € 7.

2.A.4 Practice question second part

We will now ask you some questions to check your understanding. You can always browse back to previous screens. When you have a question for the experimenter, please raise your hand.

1. Are the rules that determine how a group decision is reached different in this part compared to in part 1?

2. In a period where you can choose between individual and group decision-making, is it possible that you have to decide individually although you chose group decision-making?

3. In a period where you can choose between individual and group decision-making, is it always in your best interest to truthfully state which mode of decision-making you prefer?

4. In the next 16 periods, are there also periods in which you do not have the possibility to choose between individual and group decision-making, but are assigned to either individual or group decision-making?

5. In periods in which you choose between individual and group decision-making (type 3), does your waiting time depend on whether you choose individual or group decision-making?

2.A.5 Belief elicitation

Before you start with the next 16 periods in part 2, we would like to know what decisions you expect others to make in different situations.

		Payoff	
		Player 1	Player 2
Choice	A	10	0
	B	6	6

We will ask you how many out of 10 randomly drawn decisions are A for decisions from four different categories of participants.

Payment

You can earn money by predicting other participants' decisions well. At the end of the experiment, we will randomly choose 10 decisions in each category excluding your own decisions and compare your predictions with the actual choices. Your prediction error is how far your prediction is off. For each quest minus your prediction error times 10 cents. ion you are asked, your earnings are 1 Your payoff from each question will be added to your payoff from part 1 and part 2 at the end of the experiment.

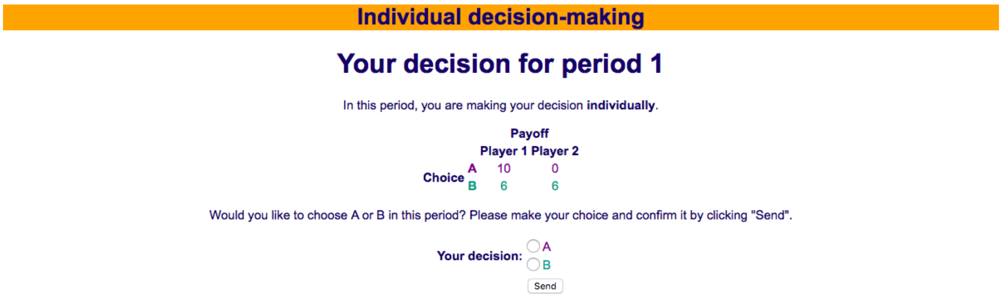
Realise that your actions in the next part will not influence your payoffs from answering these questions.

Example

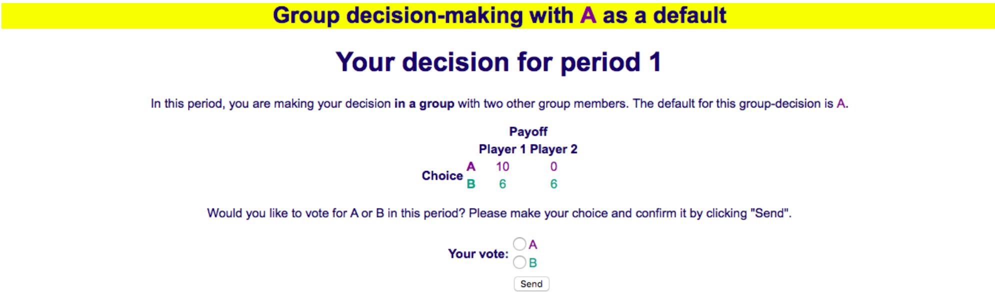
Imagine you believe that 3 out of 10 participants that prefer group decision-making if B is the default choose A. The actual number of participants that choose A out of the 10 draws of participants that prefer group decision making if B is the default is 4. Therefore, you receive as your payoff from this decision 90 cents:

$$1-0.1\times|3-4|=0.9$$

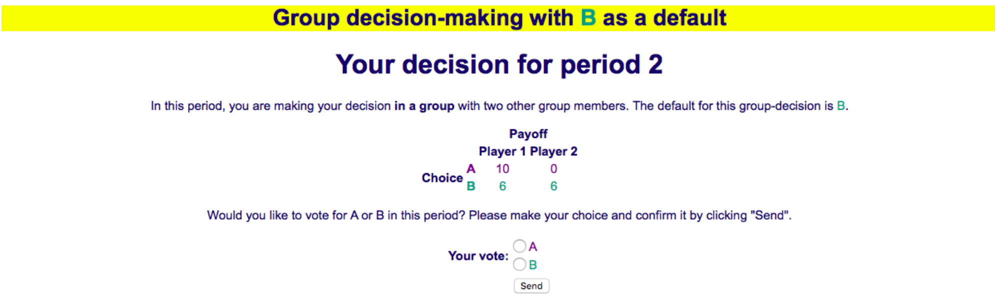
2.B Images of decision screens



(a) Individual decisions

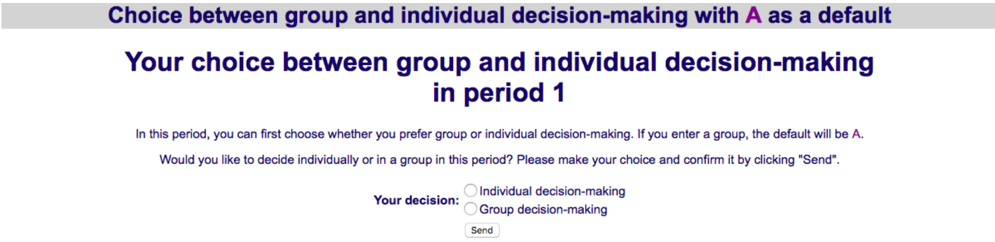


(b) Exogenous group decisions with *A* as default

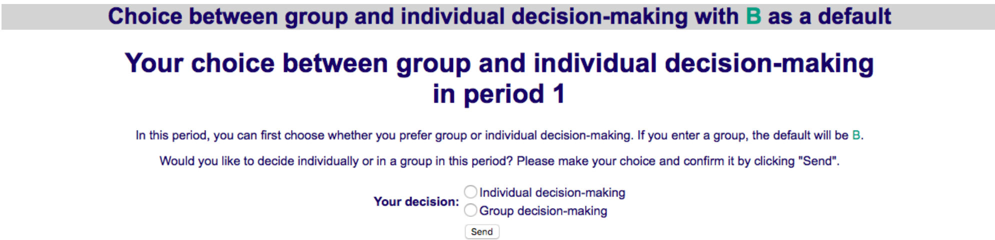


(c) Exogenous group decisions with *B* as default

Figure 2.B.1: Decision screens: Exogenous periods

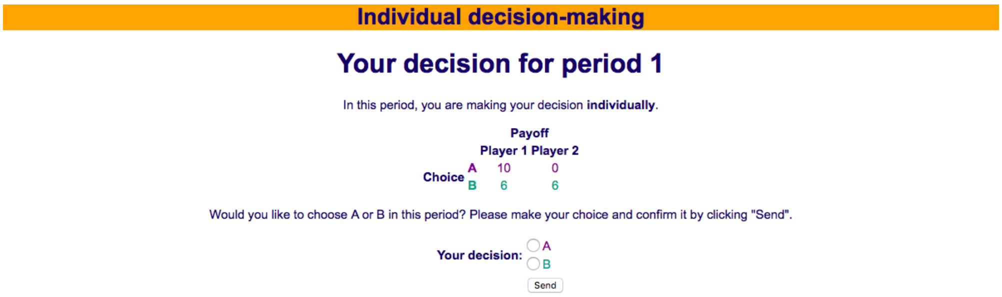


(a) Selection with *A* as default

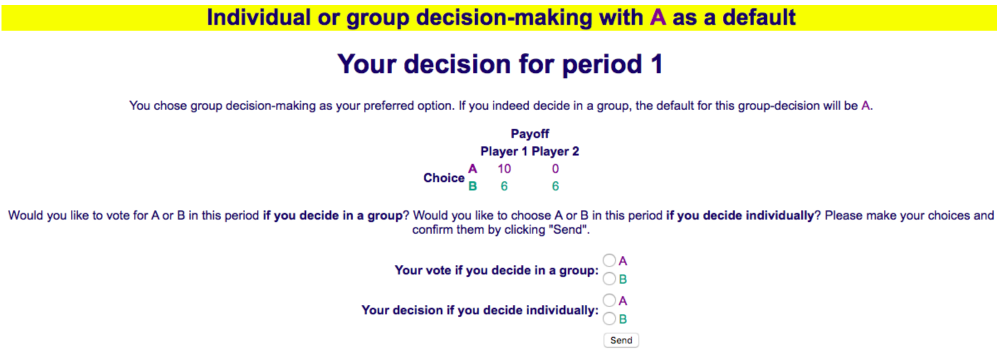


(b) Selection with *B* as default

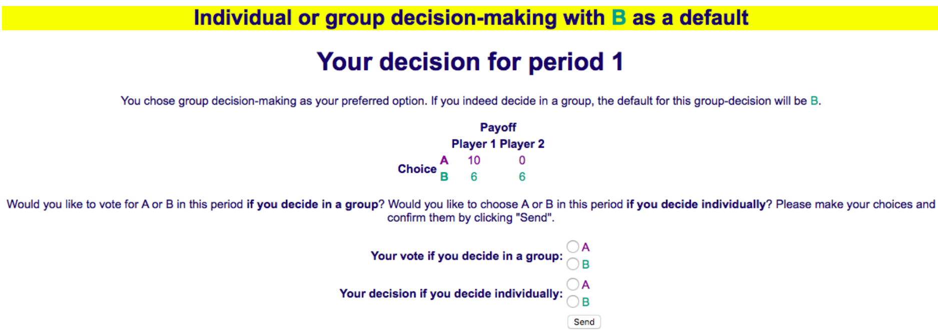
Figure 2.B.2: Decision screens: Selection in endogenous periods



(a) Individual decisions



(b) Endogenous group decisions with A as default



(c) Endogenous group decisions with B as default

Figure 2.B.3: Decision screens: Endogenous periods

2.C Decision-theoretic model

To formalise the intuitive reasoning about the diffusion of responsibility that underlies the hypotheses presented in the main text, we present here a model of individual decision making that yields the same hypotheses.

2.C.1 Theoretical framework and results

The model's essential ingredient is that individuals are averse to being responsible for a selfish action. We assume that individuals are heterogeneous in the degree of this responsibility aversion, which is captured by $\alpha_i \geq 0$ for individual i. Borrowing from Engl (2018), responsibility is measured as the distance to being pivotal, d_i. The distance to being pivotal is defined as the number of votes that need to change (including agent i's vote) such that agent i is pivotal for changing the outcome. The model adopts a consequentialist approach as in Rothenhäusler et al. (2018) by assuming that an agent only feels responsible if voting in favour of taking a selfish action and this outcome materialises. We further assume that there is a monotone linear relationship between the reduction of the other party's monetary payoff and the subsequent decline of i's utility through responsibility aversion. We are interested in the dictator's decision and denote the dictator as Player 1 and the receiver as Player 2. We denote by i a Player 1, who may be in a group with other dictators. Her utility U_i^A from supporting the selfish action A can then be defined as

$$U_i^A(\alpha_i, d_1, \boldsymbol{\pi}) = \pi^B + \mathbb{1}_{\{Y_{-i} \geq k-1\}} \left(\pi^{A1} - \pi^B - \frac{\alpha_i}{d_i} \left(\pi^B - \pi^{A2} \right) \right)$$

Here, k denotes the number of votes needed to select the selfish action A and Y_{-i} denotes the number of other dictators in the group that choose that action. π refers to an option's monetary payoff, with π^B referring to the payoff of choosing B, which is the same for both players. π^{A1} and π^{A2} denote the payoff from choosing A for Player 1 and Player 2, respectively. $\mathbb{1}$ is the indicator function.

The realised utility U_i^B from supporting B is given by

$$U_i^B(\alpha_i, d_i, \boldsymbol{\pi}) = \pi^B + \mathbb{1}_{\{Y_{-i} \geq k\}} \left(\pi^{A1} - \pi^B - \alpha_i c_i \left(\pi^B - \pi^{A2} \right) \right)$$

$\alpha_i c_i$ captures i's (non-diffusive) moral costs associated with being part of a group that implements a selfish choice without i voting in favour of it. This is modelled to be a fraction of the responsibility aversion.

We assume $\frac{\mathrm{E}\left(\frac{1}{d_i}\right)}{1+p_0} < c_i < \mathrm{E}\left(\frac{1}{d_i}\right)$, where p_j is the subjective probability that i attributes to j other agents in her group voting for A. $\mathrm{E}\left(\frac{1}{d_i}\right)$ is i's expected responsibility when voting selfishly, which is given by $p_0 + p_1\frac{1}{2} + p_2\frac{1}{3}$. The expected costs associated with voting selfishly when implementing A provide a natural upper bound on the moral costs of being part of a group that implements A while not voting selfishly. The assumption on the lower bound of c_i ensures that the model truly captures the diffusion of responsibility. It can be interpreted as c_i being high enough to prevent agents who vote B from free-riding on others' A choices. If i believes that A is likely to be implemented in any case, she still does not benefit from joining a group and voting for B.[18]

Furthermore, we assume that in case of indifference individuals prefer B over A. This can be motivated by Fehr and Schmidt (1999) or Charness and Rabin (2002) type of preferences: If individuals are indifferent, they prefer the equitable and efficient allocation.

If decisions are made individually, $k = 1$ and $d_i = 1$. For group decisions where B is the default, $k = 3$ and, given the lack of any possibility to of responsibility being diffused, $d_i = 1$. If A is the default $k = 1$ and $d_i = Y_{-i}$. Finally, we denote by $\omega = \frac{\pi^{A1}-\pi^{B}}{\pi^{B}-\pi^{A2}}$ the advantage that the dictator has relative to the receiver's loss when B is replaced by A. The dictator gains more (less) than the receiver loses if $\omega > (<)1$. For our experimental parameters, $\omega = \frac{2}{3}$, because the move from B to A improves the dictator's earnings by 4 and reduces the receiver's by 6.

This setup allows us to derive the following propositions.

Proposition 1. *Individual decision makers choose A if and only if $\alpha_i < \omega$.*

Proof. As individuals choose B if they are indifferent, individual i chooses A if and only if

$$\pi^{A1} - \alpha_i\left(\pi^{B} - \pi^{A2}\right) > \pi^{B} \Rightarrow \alpha_i < \frac{\pi^{A1} - \pi^{B}}{\pi^{B} - \pi^{A2}} = \omega$$

□

Only individuals not too averse to being responsible for reducing someone else's payoffs choose the selfish option. If the monetary benefits of taking the selfish action $(\pi^{A1} - \pi^{B})$ are larger than the responsibility felt for the dictator's loss $\alpha_i(\pi^{B} - \pi^{A2})$, then the dictator will vote for A.

Proposition 2. *In groups with B as the default, individuals vote for A if and only if $\alpha_i < \omega$.*

Proof. With individuals voting in favour of B if they are indifferent, individual i votes

[18]The costs associated with this action are $(p_1 + p_2)\alpha_i c_i\left(\pi^{B} - \pi^{A2}\right)$. These costs are higher than the gains $(p_1 + p_2)\left(\pi^{A1} - \pi^{B}\right)$ for the individuals concerned.

for A if B is the default if and only if

$$\mathrm{E}\left(\pi^B + \mathbb{1}_{\{Y_{-i}\geq k-1\}}\left(\pi^{A1} - \pi^B - \frac{\alpha_i}{d_i}\left(\pi^B - \pi^{A2}\right)\right)|i \text{ votes for } A\right) >$$
$$\mathrm{E}\left(\pi^B + \mathbb{1}_{\{Y_{-i}\geq k\}}\left(\pi^{A1} - \pi^B - \alpha_i c_i\left(\pi^B - \pi^{A2}\right)\right)|i \text{ votes for } B\right).$$

Because B is the default, $k = 3$. On the l.h.s. of the inequality, A is implemented if $Y_{-i} \geq k-1$ and therefore $d_i = 1$. The inequality then reduces to

$$\mathrm{E}\left(\mathbb{1}_{\{Y_{-i}\geq 2\}}\left(\pi^{A1} - \pi^B - \alpha_i\left(\pi^B - \pi^{A2}\right)|i \text{ votes for } A\right)\right) > 0$$

This is satisfied if and only if

$$\pi^{A1} - \pi^B - \alpha_i\left(\pi^B - \pi^{A2}\right) > 0 \Rightarrow \alpha_i < \frac{\pi^{A1} - \pi^B}{\pi^B - \pi^{A2}} = \omega.$$

□

Intuitively, since individuals also bear the full responsibility for selfish choices when deciding in a group with B as the default, the same threshold α_i characterises the set of individuals voting A in groups with B as the default as the set choosing A in the individual decision.

Proposition 3. *In groups with A as the default, individuals vote for A if and only if* $\alpha_i < \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)-c_i}\omega$ *for* $p_0 \neq 0$.

Proof. Assuming that individuals vote for B if they are indifferent, individual i votes for A if A is the default if and only if

$$\mathrm{E}\left(\pi^B + \mathbb{1}_{\{Y_{-i}\geq k-1\}}\left(\pi^{A1} - \pi^B - \frac{\alpha_i}{d_i}\left(\pi^B - \pi^{A2}\right)\right)|i \text{ votes for } A\right) >$$
$$\mathrm{E}\left(\pi^B + \mathbb{1}_{\{Y_{-i}\geq k\}}\left(\pi^{A1} - \pi^B - \alpha_i c_i\left(\pi^B - \pi^{A2}\right)\right)|i \text{ votes for } B\right)$$

Because $k = 1$ when A is the default, this reduces to

$$\pi^{A1} - \pi^B - \mathrm{E}\left(\frac{1}{d_i}\right)\alpha_i\left(\pi^B - \pi^{A2}\right) > (p_1 + p_2)\left(\left(\pi^{A1} - \pi^B\right) - \alpha_i c_i\left(\pi^B - \pi^{A2}\right)\right)$$
$$\Rightarrow \alpha_i < \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - c_i}\omega$$

This uses $c_i < \mathrm{E}\left(\frac{1}{d_i}\right)$, $1 - p_1 - p_2 = p_0$, and $p_0 \neq 0$. □

Compared to the case with B as a default, i's expected degree of responsibility enters the threshold for α_i and inflates this threshold, because $\frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)-c_i} \geq 1$.[19] Therefore,

[19]Since $c_i > \frac{\mathrm{E}\left(\frac{1}{d_i}\right)}{1+p_0}$, $\frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)-c_i} \geq \frac{1+p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)} \geq 1$ for $p_0 \neq 0$. This uses $E\left(\frac{1}{d_i}\right) = p_0 + p_1\frac{1}{2} + p_2\frac{1}{3} \leq p_0 + 1$.

some levels of responsibility aversion will make individuals vote for A in groups with A as the default but for B when B is the default. For this reason, default A is predicted to increase the frequency of selfish choices compared to default B. In this way, the model predicts Hypothesis 1 of the main text.

Next, we consider the environment with endogenous group formation. An individual will choose to join a group if the expected utility from the group decision exceeds the expected utility from individual decision making.

Proposition 4. *If B is the group's default, individuals with $\alpha_i < \omega$ will never join a group; other individuals are indifferent between joining a group and deciding individually.*

Proof. i joins a group if the expected utility from doing so, $EU(\text{Group})$, is higher than the expected utility from making the decision individually $EU(\text{Ind})$.

If $\alpha_i < \omega$ and B is the default, i both chooses A individually (Proposition 1) and votes for A in the group (Proposition 2). Therefore, i joining a group requires

$$EU(\text{Group}) \geq EU(\text{Ind}) \iff (p_0 + p_1)\pi^B + p_2\left(\pi^{A1} - \alpha_i\left(\pi^B - \pi^{A2}\right)\right) \geq \pi^{A1} - \alpha_i\left(\pi^B - \pi^{A2}\right) \iff (1-p_2)\pi^B \geq (1-p_2)\left(\pi^{A1} - \alpha_i\left(\pi^B - \pi^{A2}\right)\right)$$

Now note that $\alpha_i < \omega$, implies $\pi^{A1} - \alpha_i\left(\pi^B - \pi^{A2}\right) > \pi^B$, so EU(Ind)< EU(Group). i prefers individual decision making.

If $\alpha_i \geq \omega$ and B is the default, i chooses B individually (Proposition 1) and votes for B in the group (Proposition 2). Therefore, i strictly preferring group decision making requires that

$$EU(\text{Group}) = \pi^B > \pi^B = EU(\text{Ind})$$

which is never satisfied. In this case, i is indifferent between individual and group decision making. □

Agents with $\alpha_i < \omega$ refrain from joining a group because their preferred option A may not be implemented and, since there is no mechanism through which responsibility can be diffused when the default is B, these agents have nothing to gain from joining a group. Individuals with $\alpha_i \geq \omega$ are indifferent between group and individual decision making, as they will vote for B in groups and therefore, given the default, in both cases receive the payoff corresponding to this option. With Fehr and Schmidt (1999) or Charness and Rabin (2002) type of preferences, individuals could however prefer to join a group to ensure the equitable share of resources for a more people.

Proposition 5. *If A is the group's default, only individuals with $\alpha_i < \frac{1}{E\left(\frac{1}{d_i}\right)}\omega$ will join a group.*

Proof. If $\alpha_i < \omega$ and A is the default, i both chooses A individually and votes for A in the group. Therefore, for i to (weakly) prefer group decision making, it is necessary and sufficient that

$$EU(\text{Group}) = \pi^{A1} - \alpha_i \mathrm{E}\left(\frac{1}{d_i}\right)\left(\pi^B - \pi^{A2}\right) \geq \pi^{A1} - \alpha_i \left(\pi^B - \pi^{A2}\right) = EU(\text{Ind})$$

which is always satisfied because $\mathrm{E}\left(\frac{1}{d_i}\right) \leq 1$. If $\omega \leq \alpha_i < \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - c_i}\omega$ and A is the default, i chooses B individually but votes for A in a group. For i to (weakly) prefer group decision making, it is necessary and sufficient that

$$EU(\text{Group}) = \pi^{A1} - \alpha_i \mathrm{E}\left(\frac{1}{d_i}\right)\left(\pi^B - \pi^{A2}\right) \geq \pi^B = EU(\text{Ind})$$

This condition can be rewritten as

$$\frac{1}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega \geq \alpha_i$$

We know that

$$\omega \leq \frac{1}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega \leq \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - c_i}\omega$$

Hence, $\frac{1}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega \geq \alpha_i$ is satisfied for individuals with $\omega \leq \alpha_i \leq \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega$, but not for those with $\frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega < \alpha_i \leq \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - c_i}\omega$. Finally, if $\frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - c_i}\omega \leq \alpha_i$ and A is the default, i chooses B individually and votes for B in a group. To prefer group decision making, it is necessary that

$$EU(\text{Group}) = p_0\pi^B + (p_1 + p_2)\left(\pi^{A1} - \alpha_i c_i \left(\pi^B - \pi^{A2}\right)\right) \geq \pi^B = EU(\text{Ind})$$

This condition can be rewritten as

$$\omega \geq \alpha_i c_i$$

Given the assumed $c_i > \frac{\mathrm{E}\left(\frac{1}{d_i}\right)}{1+p_0}$ and $\alpha_i \geq \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - c_i}\omega$, we have

$$\alpha_i c_i > \frac{\mathrm{E}\left(\frac{1}{d_i}\right)}{1 + p_0} \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right) - \frac{\mathrm{E}\left(\frac{1}{d_i}\right)}{1+p_0}}\omega = \omega$$

Hence, the condition cannot be satisfied and i in this case never wants to join a group.

□

When A is the default, agents with $\alpha_i < \omega$ have an incentive to select into group decision making, since they might benefit from the diffusion of responsibility and can ensure the high payoff from A by voting for A. Also individuals with $\omega \leq \alpha_i \leq \frac{1}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega$ join groups, because for their level of responsibility aversion, the higher payoff associated with A together with possibly diffused responsibility make group decision making more attractive. For individuals with $\alpha_i > \frac{1}{\mathrm{E}\left(\frac{1}{d_i}\right)}\omega$, selecting into group decision making is not appealing, since a potentially higher payoff cannot compensate for the moral costs associated with bearing at least part of the responsibility for a selfish outcome (those that vote for A) or being part of a group that causes this outcome (those that vote for B).

With A as the default, for individuals with $\alpha_i \leq \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)-c_i}\omega$, the attractiveness of joining a group is increasing in the expected number of individuals voting for A, since this both increases the likelihood of receiving the high payoff and the expected distance to being pivotal for agents voting for A. For agents with $\alpha_i > \frac{p_0}{\mathrm{E}\left(\frac{1}{d_i}\right)-c_i}\omega$, the opposite holds true, since the costs associated with being part of a group that implements a selfish outcome are higher than the expected monetary gains from a higher payoff.

It follows from Proposition 4, that any individual joining a group with B as the default has a high responsibility aversion. They will vote for B (Proposition 2). Proposition 5 shows that when A is the default, it is those with low responsibility aversion that join groups; they vote A (Proposition 3). Compared to the case of exogenous group formation, this amplifies the difference between the two defaults in terms of the number of A votes. This is how the model predicts the comparative statics formulated in Hypothesis 2 of the main text. Together, Propositions 4 and 5 imply that those with a low aversion to responsibility are more likely to join a group when A is the default than when B is. This is what Hypothesis 3 predicts. Finally, these propositions predict that an individual with high responsibility aversion might join a group when B is the default (they are indifferent), but will not do so A is the default. This is captured by Hypothesis 4.

2.D Traditional non-parametric tests

The following tables replicate the results from Table 2.3, 2.6, 2.4 and 2.7 using Wilcoxon signed-rank and Mann-Whitney U tests instead of permutation tests.

	A votes			Group Entry				
	Exo		Endo	Total		Selfish	Switching	Pro-social
Overall	0.372	$>$	0.3078	0.4007		0.2778	0.3894	0.4734
A default	0.3843	$<$	0.375	0.3796		0.3542	0.4038	0.3875
	$\vee$	$\wedge^{***}$	$\vee^{***}$	$\wedge^{*}$		$\vee^{**}$	$\vee$	$\wedge^{***}$
B default	0.3596	$>^{**}$	0.2473	0.4218		0.2014	0.375	0.5593
				Others				
Difference				-0.0812	$<^{***}$	0.1528		
between	0.0247	$<^{***}$	0.1277	-0.0647	$<^{**}$		0.0288	
defaults				0.0343	$>^{***}$			-0.1718
Observations	692		863	1727		288	416	640

Notes: Average fraction of *A* votes in groups and average fraction of individuals choosing to enter a group. Differences of means are all tested with Wilcoxon signed-rank tests using matching group averages, except for differences for different types, which are tested with Mann-Whitney U tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.D.1: Summary statistics with traditional non-parametric tests

	A votes					
	Only $1st$ part			$1st$ and $2nd$ part		
	Exo Group		Exo Ind	Exo Group		Exo Ind
A default	0.44	$<$		0.41	$<$	
	$\vee^{*}$		0.48	$\vee^{**}$		0.44
B default	0.38	$<^{**}$		0.37	$<^{***}$	
Observations	648		648	1080		1513

Notes: Cells show per default the average fraction of *A* votes in exogenously formed groups and exogenously imposed individual decision making. Differences in means are tested with Wilcoxon signed-rank tests using matching group averages. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table 2.D.2: Comparison of votes including first part of the experiment

	fraction of type in group		
	Selfish	Switching	Pro-social
A default	0.16	0.26	0.38
	$\vee^{**}$	$\vee$	$\wedge^{***}$
B default	0.08	0.21	0.49
Groups	24	24	24

Notes: Cells report the fraction of group members that have a certain type in groups with *A* and *B* as the default. The unit of observation is the average per matching group. Differences are tested with Wilcoxon signed-rank tests using matching group averages. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.D.3: Composition of endogenously formed groups with traditional non-parametric tests

	A votes					
	Endo Group		Endo Ind	Exo Group		Exo Ind
A default	0.375	$<$	0.3918	0.3843	$<$	
	$\vee^{***}$		$\wedge^{***}$	$\vee$		0.4104
B default	0.2473	$<^{***}$	0.479	0.3596	$<$	
Observations	692		1035	863		865

Notes: Average fraction of *A* votes in groups compared to average fraction of *A* votes of individuals in either endogenous or exogenous group formation periods. Differences of means are tested using matching group averages with Wilcoxon signed-rank tests, except for differences between individuals and groups in endogenous periods, which are tested with Mann-Whitney U tests. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.D.4: Comparison of group and individual decisions with non-parametric tests

2.E Different type classification mechanisms

To test the robustness of our findings regarding type-specific behaviour, we examine different classification mechanisms.

First, we repeat the analysis while classifying types based on their relative selfishness. We focus on the classification of selfish and pro-social types, not switching types. As our hypotheses only concern these two types, this does not impose a restriction. Table 2.E.1 summarises the results, using the fraction of selfish choices in exogenous group decisions as a measure. Table 2.E.2 uses the fraction of selfish choices in all exogenous periods, including individual decision-making periods, instead.

	Group Entry			
	(1)	(2)	(3)	(4)
Relative selfishness	Top 25%	Bottom 25%	Top 50%	Bottom 50%
A Default	0.375	0.3875	0.3763	0.3821
	$\vee^{**}$	$\wedge^{***}$	$\vee^{**}$	$\wedge^{**}$
B Default	0.2902	0.5593	0.3011	0.5132
Observations	448	640	744	983

Notes: Average fraction of individuals choosing to enter a group. Column (1) considers individuals whose fraction of selfish choices is within the upper quartile, column (2) individuals whose fraction of selfish choices within the lower quartile, column (3) individuals whose fraction of selfish choices is not below the median, column (4) individuals whose fraction of selfish choices is below the median. For this classification, only decisions in exogenously formed groups in the second part of the experiment are considered. Differences of means are all tested with Fisher-Pitman permutation tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.1: Robustness check of type-specific behaviour: Classification based on decisions in exogenously formed groups

	Group Entry			
Relative selfishness	(1) Top 25%	(2) Bottom 25%	(3) Top 50%	(4) Bottom 50%
A Default	0.3538	0.307	0.4414	0.3143
	$\vee^{*}$	$\wedge^{***}$	$\vee^{**}$	$\wedge^{**}$
B Default	0.2625	0.5395	0.386	0.4595
Observations	519	456	887	840

Notes: Average fraction of individuals choosing to enter a group. Column (1) considers individuals whose fraction of selfish choices is within the upper quartile, column (2) individuals whose fraction of selfish choices within the lower quartile, column (3) individuals whose fraction of selfish choices is not below the median, column (4) individuals whose fraction of selfish choices is below the median. For this classification, decisions in all exogenous periods in the second part of the experiment are considered. Differences of means are all tested with Fisher-Pitman permutation tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.E.2: Robustness check of type-specific behaviour: Classification based on decisions in all exogenous periods

2.F Beliefs and selection choices

Sample	(1) Total	(2) Selfish	(3) Switching	(4) Pro-social
	Entry	Entry	Entry	Entry
Belief on *A* votes	-0.00867	0.111	0.0596	-0.107
	(0.0534)	(0.121)	(0.0895)	(0.0872)
	[0.8949]	[0.3774]	[0.5085]	[0.2342]
Constant	-0.453	-1.364	-0.721	-0.00718
	(0.332)	(0.885)	(0.537)	(0.479)
Observations	860	144	208	316
Clusters	24	17	23	21

Notes: Logit regression estimating the effect of individuals' beliefs on the number of *A* votes in groups with *A* as a default on selecting into groups. Standard errors clustered at the matching group level are in parentheses. In square brackets, score wild cluster bootstrap *p*-values adjusted for small samples are reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.F.1: The effect of beliefs on selection into groups

CHAPTER

3

HOW TEAMS OVERCOME FREE-RIDING IN STRATEGIC EXPERIMENTATION

3.1 Introduction

Innovation is at the core of overcoming today's most pressing issues, from fighting climate change to tackling the COVID-19 pandemic. Estimates by the International Energy Agency suggest that almost half of the required emission reductions on the road to net-zero emissions by 2050 will require technologies that are not yet available (Bouckaert et al., 2021). During the height of the pandemic, scientific teams worldwide worked on better treatments for COVID-19 and created effective vaccines to combat it. Teams in companies, governmental agencies, and civil organisations are trying to find innovative approaches that help the economy, improve communal lives in times of social distancing and keep people healthy (see e.g. Kretchmer, 2020).

Experimentation, that is testing new technologies with uncertain outcomes, is central to innovation (Thomke, 2003). Innovative team projects often require a process of trial and error with the risk of spending time and resources on a doomed project. Consider a group of employees working on a business innovation or researchers developing a new drug or technology. As an illustrating example from the medical domain, think of a team in a pharmaceutical company attempting to develop a new vaccine (i.e., the team is *experimenting*). Not all attempts will lead to success, and some will entail a waste of resources, which could have been spent on other projects with a certain reward. Recent estimates suggest that only 13.8% of all drug development programs resulted in FDA approval (Wong et al., 2019).

This project studies how the experimentation environment can be designed to encourage experimentation in teams. Large innovations usually require a team of individuals to experiment together, such as teams within companies or governmental organisations, but also scientific collaborations. Over the last decades, such collaborations have substantially increased (Dong et al., 2017). If teams are experimenting, individual contributions to new projects may provide a public good to all group members. In the illustrating example, if one team member observes that a certain vaccine indeed is effective (i.e., there is a *breakthrough*), all individuals in the team may benefit from the completion of their joint task (i.e., the breakthrough is a *public good*). Crucially, this public good entails two dimensions: The information created from observing successful experimentation, and the payoffs benefiting all team members.

This can result in a two-dimensional free-riding problem. First, there is the well-known moral hazard in teams problem, going back to Holmstrom (1982). Agents prefer that other team members invest their resources, such as time or individual budgets, in a project compared to investing themselves if the payoff of a breakthrough is shared among all team members. This causes lower experimentation investment than socially optimal. Aside from free riding on others' experimentation investment due to the payoff externality, informational spillovers change the incentives for experimentation. Acquiring information about a project's quality yourself, such as conducting a trial of the effectiveness of a vaccine, is costly, so individuals prefer to use the information generated by others. Hence, if the information created through experimentation is public, there is an informational externality that results in a free-riding problem. This also leads to inefficiently low experimentation efforts (Bonatti and Hörner, 2011). In this study, I investigate in the laboratory which types of experimentation environments allow agents to overcome these free-rider problems.

Informational spillovers pose challenges for individuals experimenting in a team. Individuals need to anticipate how their fellow team members will react to the information generated by their own experimentation efforts. At the same time, they also have to carefully consider what other team members' actions reveal about the quality of the project they are engaged with and respond to this information. In this paper, I will empirically examine how individuals handle these challenges. In doing so, this paper will study whether and how teams can overcome free-riding problems when experimenting with projects of uncertain quality. I will focus on the dimensions of experimentation that are specifically relevant to teams.

The first crucial aspect of experimentation in teams is that how agents learn from the experimentation of team members depends on whether the actions of team members are observable for all team members. Distinct settings vary in the observability of experimentation effort. Some teams may find it easier than others to observe their team member's input in a group project. These settings are theoretically well understood, but less so empirically. It is thus important to gain a better understanding of how the provision of information on experimentation efforts changes behaviour. At the same time, there are differences in the extent to which one team member's success is predictive of another team member's likelihood of success. This is the next aspect under consideration.

Specifically, I will study settings where agents either work on separate, independent projects or jointly work on one project. In the former case, the success of one team member does not provide any information for others. Going back to the leading example, the individual team members may explore distinct technologies to develop a vaccine to combat a certain disease. The latter case, with one joint project, is the polar opposite. Here, the experimentation of all team members is equally informative, and informational spillovers are a natural part of the environment. In the example, joint experimentation implies that all team members work on one specific vaccine, relying on the same technology.

The experimental design builds on a simple theoretical model, close to an example in Bonatti and Hörner (2011). I employ a two-stage variant with two agents to focus on how individuals utilize information provided by others and how they take into account the information they themselves generate. A breakthrough is a public good and results in a positive payoff for all team members. In this model, a breakthrough reveals that the project is of high quality. If no breakthrough occurs, agents can continue experimenting, but should realize that the the project is less likely to be of high quality.

The model allows for two observations when considering joint experimentation with a common project. First, the experimentation optimal effort will be inefficiently low, because agents do not internalise the positive externalities they have on others when choosing their effort. Second, an agent's current effort choice and the other agent's future effort choice are strategic substitutes. This implies that, if effort is observable, high effort levels are unattractive, since the other agent will interpret the fact that no breakthrough occured as a strong negative signal that makes her more pessimistic about the quality of the project, reducing future effort provision. As agents anticipate this force when choosing early experimentation, they are discouraged from choosing high experimentation levels, the so-called discouragement effect. To minimise inefficiencies, it is thus not desirable that effort is observable.

Contrary to this theoretical channel, several behavioural factors would suggest that the observability of experimentation will not result in lower levels of experimentation. This paper will systematically study these factors. First, the discouragement effect hinges on agents updating their beliefs in response to the information created by others, which agents may do insufficiently. Myopic behaviour can lead to agents disregarding the effect of their early experimentation on a later stage. Second, conditional cooperation or reciprocity can result in agents encouraging each others' experimentation, and punishing low experimentation, if effort is observable. Last, it is possible that the observability of experimentation efforts allows an agent to 'lead by example', signalling the belief that experimentation is lucrative.

To study the mechanisms that drive strategic experimentation, I contrast the setting of joint experimentation to one of separate experimentation. If team members experiment with separate projects, there is no informational externality. Therefore, the discouragement effect disappears. With separate experimentation, the theoretical predictions flip. Since agents cannot discourage each other from experimenting, a new effect dominates that is particular to a setting with separate projects. Successful innovation only requires one breakthrough in one project. Agents should therefore respond positively to others' high experimentation, as a more pessimistic partner reduces the likelihood of several simultaneous, and therefore inefficient, breakthroughs.

The experiment employs a 2-by-2 between-subject design that closely follows the theoretical setup, varying the observability of experimentation effort and whether the team members experiment with a joint or separate projects. In a two-stage setting, participants experiment in the first stage, allowing them to subsequently update their beliefs in response to their and others' first-stage efforts. The updated beliefs can enable them to make a more informed second-stage experimentation decision. The second stage is the final stage. Therefore, decisions and outcomes from this stage do not entail any informational value. I can therefore use the decisions in this stage to understand behaviour without informational externalities. Furthermore, I elicit beliefs about the quality of the project and the partner's effort provision to disentangle different drivers of effort provision. 384 subjects repeat the experimentation game for 30 periods each, split over the four treatments.

The experiment allows for several observations. First, there is a stark contrast to the theoretical prediction that the observability of experimentation reduces joint experimentation efforts. Instead, observability increases joint experimentation levels. Interestingly, this cannot be traced back to a lack of sophistication in belief updating. Qualitatively, subjects respond to experimentation as predicted, both when updating their beliefs and when choosing their future actions. Beliefs are updated conservatively, but in the expected direction. In addition, second-order beliefs are consistent with subjects even anticipating this response from others. Thus, there is a discouragement effect. Nevertheless, this does not lead to agents reducing their early experimentation as a response. Hence, agents behave partially myopic, not considering how their actions will impact payoffs from later experimentation. At the same time, there is no convincing evidence pointing at reciprocity as a driver of higher experimentation levels if these are observable. Other than through a change in beliefs, first-stage experimentation does not impact the partner's second-stage experimentation, which would have indicated conditionally cooperative behaviour. Second, experimentation efforts are considerably higher if individuals experiment with a joint project than with separate projects. There is no detectable difference in the response to observable effort compared to the case of joint experimentation. Therefore, the advantage of observable experimentation exists irrespectively of the presence of an informational externality.

The two factors that increase experimentation are thus 1) experimenting jointly and 2) experimentation being observable. These factors share that they increase the salience of group membership. If agents observe their partner's action, they are repeatedly made aware that they are not working on their own. Similarly, if agents work on the same project, this shares more noticeable elements of team productions. The observed patterns are, moreover, in line with agents creating norms of high experimentation and agents leading by example to foster such norms. Agents respond to high experimentation by others in later rounds, and the variance of experimentation levels is reduced if these are observable.

The remainder of this paper proceeds as follows: Section 3.2 will give a brief overview of the related literature. Section 3.3 outlines the theoretical model that underlies the experimental design and gives its predictions, Section 3.4 provides the experimental design. Finally, Section 3.5 discusses the experimental results and Section 3.6 concludes.

3.2 Related literature

Experimentation in general, and strategic experimentation in particular, are theoretically well understood. The first bandit models of experimentation go back to Bolton and Harris (1999). Hörner and Skrzypacz (2017) give an overview of the core models in the strategic experimentation literature. These models focus on the trade-off between experimentation and exploitation in a continuous-time setting. An agent faces slot machines, so-called bandits, with uncertain payoffs. In models of strategic experimentation, several agents face the same bandits. Players can learn about the underlying payoff processes by observing the outcomes of their own and others' experimentation. Arriving news comes either as breakthroughs (Keller et al., 2005), or as breakdowns (Keller and Rady, 2015). Breakthroughs have positive, breakdowns negative payoff consequences. Breakthroughs and breakdowns usually provide conclusive evidence about the payoff process, which I will also focus on, with some exceptions (e.g. Keller and Rady, 2010).[1]

The study by Bonatti and Hörner (2011) is closest to the setup studied in this paper. Bonatti and Hörner (2011) focus on experimentation where breakthroughs provide a public good. In this environment, informational and payoff externalities co-exist, creating the two-dimensional free-riding problem discussed above. Theoretically, this setup induces both free-riding and delay of experimentation. Furthermore, monitoring the other agent by observing their experimentation does not reduce delay or free-riding, as this would imply that agents discourage each other from experimenting. I will provide an experimental test of this conclusion. Adding to this, I will theoretically and experimentally contrast the case of joint experimentation, studied in Bonatti and Hörner (2011), to the case where team members experiment with separate projects.

So far, experimental tests of the theoretical predictions are scarce and primarily focus on individual, not strategic experimentation. In the laboratory, agents frequently undervalue experimentation when facing individual bandit problems (Meyer and Shi, 1995), which can be driven by risk aversion (Hudja and Woods, 2021) or ambiguity attitudes (Anderson, 2012).[2]

The theoretical predictions of the strategic experimentation literature have so far not widely been tested, though there are some notable exceptions. In a test of the model of Keller et al. (2005), there is substantial free-riding on others' experimentation (Hoelzemann and Klein, 2021). Consistent with Markov Perfect Equilibria, participants use non-cutoff strategies when experimenting (Hoelzemann and Klein, 2021). Experimentation in groups can be sustained at more pessimistic beliefs than

[1] For a discussion and comparison of the theoretical properties of these models, see Hörner and Skrzypacz (2017).

[2] There is mixed evidence on whether participants respond to parameter changes, such as changes of the discount rate and prior beliefs, in their experimentation efforts (Banks et al., 1997; Hudja and Woods, 2021).

theoretically predicted in this setting (Kwon, 2020). This implies that groups generate more information than individuals, which is in contrast to theoretical predictions. Also in a simpler two-stage setup, as employed in this paper, free-riding on others information provision exists and participants under-experiment compared to theoretical predictions (Boyce et al., 2016).

So far, there is no evidence on the influence of treatment variations that would allow drawing conclusions on how distinct elements of strategic experimentation, such as the existence of an informational externality or whether experimentation is observable, impact experimentation efforts. In contrast to existing studies, I test the comparative statics of how behaviour depends on the observability of experimentation effort in a setting where payoff externalities exist. I do not aim at giving insights into the dynamics of behaviour in a continuous-time setting; instead, I employ a simpler discrete-time setting and focus on the determinants of experimentation. Furthermore, I am interested in how potential biases in belief formation drive experimentation, which could not be clearly studied in earlier work, where belief updating was trivial (Kwon, 2020).

There are recent experimental and theoretical papers that look at collaborative search. Search differs from strategic experimentation and the setup that is studied in this paper, because agents are not exploring the merits of one particular policy or technology but explore a set of such items. The overarching questions, nevertheless, are similar. Both collaborative search and strategic experimentation look at how teams can provide innovation. In collaborative search, however, agents encourage each others' search, as a breakthrough becomes more likely when more projects have been examined. In a setting of collaborative search with payoff externalities, imperfect optimisation and other-regarding preferences influence agents' experimentation (von Essen et al., 2020).

My research also relates to the public good literature. While this study does not look at a deterministic public good, experimenting increases the probability of a breakthrough, which constitutes a payoff public good. Therefore, findings in the public goods literature could help us to understand the behavioural drivers of experimentation. In the experimental public good literature, many people are conditional cooperators and match contributions by others (see e.g. Fischbacher et al., 2001; Kocher et al., 2008; Thöni and Volk, 2018; Croson et al., 2005). The presence of conditional cooperators would imply that agents experiment more if they see others experiment more as well. Reciprocal behaviour allows for the provision of public goods, even if that is not in the agents' direct economic interests (see e.g. Sugden, 1984).

In public good games, group membership plays an important role. The salience of group membership increases the weight that agents put on payoffs for their group members (Charness et al., 2007; Sutter, 2009). Changing the observability of others' actions and whether partners work on the same project will likely also increase the salience of group membership and could therefore increase experimentation in this experiment.

Similarly, observable experimentation may provide incentives to 'lead by example', as observed in public good experiments (Vesterlund, 2003; Potters et al., 2005, 2007; Güth et al., 2007; Levati et al., 2007). If agents observe each others' experimentation, it may prove beneficial to set high levels of experimentation to encourage future experimentation by others, either by signalling the profitability of experimentation, or by creating norms of high experimentation.

3.3 Theoretical framework

In this section, I will introduce the theoretical model underlying the experimental design. This model provides the theoretical predictions tested in the experiment, illustrating the drivers of strategic experimentation. The theoretical framework builds on a variant of a two-stage model based on Bonatti and Hörner (2011). Applying a simpler setting allows me to focus on how individuals create information and utilize the information provided by others and by themselves. I study a model where a breakthrough is a public good. Therefore, everyone in a team receives a positive payoff if a team member achieves a breakthrough. News is always good, as breakthroughs reveal that the project is of high quality. If no breakthrough occurs, agents can continue experimenting.

In this two-stage model, each stage facilitates the analysis of a distinct element of strategic experimentation. First-stage experimentation captures that agents generate new information and that their experimentation entails an informational externality. This is the core element of experimentation. Second-stage experimentation captures the response to the information previously created. As the second stage is the final stage, no informational value can be generated. Therefore, there is also no informational externality.

First, I consider a setting where teams experiment jointly with one project. In a second step, I adapt this model to encompass teams in which team members experiment with separate projects to achieve a breakthrough. For both cases, I will differentiate between a setting where experimentation efforts are observable to the other team member and a setting where these efforts are not observable.

3.3.1 Experimenting with a joint project

This two-stage model studies joint experimentation. There are two agents $i = 1, 2$ who can choose to invest experimentation effort $e_{i,t} \in [0, 1]$ in two stages $t = 1, 2$ in a joint project with unknown quality. Doing so entails a private cost of effort of $c(e_{i,t}) = 2e_{i,t}^2$. Both agents receive a payoff of $Y = 13$ from the project if a breakthrough occurs. A breakthrough terminates the project. Whether a breakthrough occurs depends on the quality of the project, which can be high or low, and on the effort the two agents invest in that project. The common prior that the project is of high quality is $p = 0.5$. Conditional on the project being of high quality, the probability that a breakthrough occurs in stage t is given by $\frac{e_{i,t}+e_{-i,t}}{2}$, which is increasing in the effort invested by both agents. If the project is of low quality, there will never be a breakthrough.

Setting $Y > 2c'(e_{i,t})$ here ensures that if a project's quality would be known and high, agents would find the reward of experimenting promising enough to invest full effort in the project, as the marginal benefit is larger than the marginal costs for all $e_{i,t}$.

The experimentation effort by the two team members is either observable or not. I first consider the case in which agents observe their team member's level of first-stage experimentation before choosing their second-stage experimentation levels. In the second stage of the experiment, agents maximise the following expected utility:

$$EU_{i,2} = \rho(e_{i,1}, e_{-i,1})\left(\frac{e_{i,2}+e_{-i,2}}{2}\right)Y - c(e_{i,2}) \tag{3.1}$$

This stage is only reached if there was no breakthrough. $\rho(e_{i,1}, e_{-i,1})$ is the posterior belief that the project is of high quality. By Bayes' rule, this is given by

$$\rho(e_{i,1}, e_{-i,1}) = \frac{p\left(1-\frac{e_{i,1}+e_{-i,1}}{2}\right)}{1-p\left(\frac{e_{i,1}+e_{-i,1}}{2}\right)} \leq p$$

Here, realise that the posterior is decreasing both in the agent's own and in their partner's first-stage experimentation:

$$\frac{\partial \rho(e_{i,1}, e_{-i,1})}{\partial e_{-i,1}} \leq 0 \quad \text{and} \quad \frac{\partial \rho(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} \leq 0, \qquad \forall p \in [0,1]$$

Intuitively, if there is no breakthrough but experimentation efforts are high, it is less likely that the project is of high quality. If the project were of high quality, a breakthrough would have been likely.

In the first stage, agents maximise the expected utility over the two stages, taking into account how first-stage experimentation affects second-stage experimentation. The first-stage expected utility as a function of the second-stage expected utility $EU_{i,2}$ is given by

$$EU_{i,1} = p\left(\frac{e_{i,1}+e_{-i,1}}{2}\right)Y - c(e_{i,1}) + \left(1-p\left(\frac{e_{i,1}+e_{-i,1}}{2}\right)\right) \times EU_{i,2} \tag{3.2}$$

I will consider experimentation behaviour in the pure-strategy symmetric Perfect Bayesian Nash equilibrium (PBE). Similar to Bonatti and Hörner (2011), this setup allows us to make the following observations:

Lemma 1. *An agent's first- and second-stage experimentation efforts are strategic substitutes.*

An agent's second-stage effort is an increasing function of the posterior belief of the project's quality $\rho(e_{i,1}, e_{-i,1})$. Second-stage experimentation promises to more likely pay off if agents are optimistic about the project's quality. The posterior $\rho(e_{i,1}, e_{-i,1})$ is decreasing in the first-stage experimentation. Therefore, the second-stage experimentation is decreasing in first-stage experimentation. Intuitively, agents become more pessimistic about the project's quality if they exerted high experimentation efforts in the first stage but did not observe a breakthrough. As a consequence, they exert less effort in the second stage.

The same underlying reasoning applies to the strategic substitutability of own experimentation and partner's experimentation across stages.

Lemma 2. *An agent's and her partner's first- and second-stage experimentation efforts are strategic substitutes.*

As the posterior belief of the project's quality $\rho(e_{i,1}, e_{-i,1})$ is also decreasing in the partner's first-stage experimentation, agents also grow increasingly more pessimistic about the project's quality the higher is their partner's first-stage experimentation.

From the two preceding lemmas, the main proposition that this model allows for follows. This proposition concerns behaviour in the PBE of the game, comparing the setting that was outlined with observable effort to a setting with unobservable effort.

Proposition 1. *If experimentation effort is unobservable, experimentation effort is higher in the first stage.*

As a breakthrough provides a public good, agents benefit from their partner's experimentation. Therefore, they want to encourage their partner to experiment in the future. Hence, agents will take into account how their action changes the posterior belief and thereby the second-stage experimentation of their partner. An agent's partner's second-stage experimentation effort is decreasing in that agent's first-stage experimentation, see Lemma 2. Thus, if agents observe each others' experimentation, every agent has an incentive to decrease their experimentation in the first stage to encourage future experimentation of their partner.

If, however, experimentation effort is not observable, agent i only forms a belief, $\hat{e}_{-i,1}$, about their partner's first stage experimentation. The posterior that enters the second-stage expected utility in Equation 3.1 is then a function of $\hat{e}_{-i,1}$ and not of $e_{-i,1}$.

Importantly, conditional on there being no breakthrough, the actual level of experimentation effort $e_{-i,1}$ has no influence on $\hat{e}_{-i,1}$ if it is unobservable. Agents know that their partner will likewise only form a belief, $\hat{e}_{-i,1}$, about their first stage experimentation. So compared to the equilibrium level of experimentation if effort is observable, agents can deviate to a higher experimentation level if this is unobservable. This increases the chance of a breakthrough in the first period without making the partner more pessimistic.

Proposition 2. *If experimentation effort is observable, experimentation effort is higher in the second stage.*

This is a direct implication of Proposition 1. First-stage experimentation is higher if unobservable. Given that beliefs are correct in the PBE, agents in the second stage are more pessimistic about the project's quality if first-stage experimentation was not observable. Agent's respond to this by exerting less experimentation effort in the second stage if this was unobservable in the first stage.

All proofs for this section are presented in Appendix 3.A.

3.3.2 Experimenting with separate projects

Next, I turn to a setting where agents work on separate projects to achieve a breakthrough. As in Section 3.3.1, two agents $i = 1, 2$ can choose to invest experimentation effort $e_{i,t} \in [0, 1]$ in stages $t = 1, 2$ in a project with unknown quality. Doing so again entails a private cost of $c(e_{i,t}) = 2e_{i,t}^2$.

The crucial difference to joint experimentation is that agents work on two separate projects. Each project is independently of high quality with $p = 0.5$. Both agents receive a payoff of $Y = 13$ if there is a breakthrough in at least one of the two projects. The probability of a breakthrough in agent i's project, conditional on her project being of high quality, is $\frac{e_{i,t}}{2}$. This probability only depends on this agent's own experimentation effort. In this setup, an agent's experimentation has the same marginal impact on the probability of a breakthrough in their own project as on the joint project in Section 3.3.1. Agents again either observe or do not observe their partner's experimentation effort.

If agents observe their partner's experimentation, they maximise the following expected utility in the second stage:

$$EU_{i,2} = \left(\rho(e_{i,1})\frac{e_{i,2}}{2} + \rho(e_{-i,1})\frac{e_{-i,2}}{2} - \rho(e_{i,1})\rho(e_{-i,1})\frac{e_{i,2}e_{-i,2}}{4}\right)Y - c(e_{i,2}) \tag{3.3}$$

Experimentation efforts by one agent are not informative about the quality of the other agent's project, since higher experimentation in one project only makes a breakthrough in that one project more likely. Breakthroughs still represent a public good, because both agents receive a payoff of $Y = 13$ if at least one of them achieves a breakthrough in their project. The posterior belief that the project is of high quality $\rho(e_{i,1})$ is, by Bayes' rule, here given by

$$\rho(e_{i,1}) = \frac{p\left(1-\frac{e_{i,1}}{2}\right)}{1-p\left(\frac{e_{i,1}}{2}\right)}$$

with

$$\frac{\partial \rho(e_{i,1})}{\partial e_{i,1}} \leq 0, \qquad \forall p \in [0,1]$$

$\rho(e_{i,1})$ only depends on the agent's own experimentation, as there is no informational externality. This implies that an agent's second-stage experimentation effort is not affected by a change in beliefs about their own project's quality that is driven by their partners' first-stage experimentation.

In the first stage, agents again consider how their experimentation will affect second-stage experimentation. They maximise[3]:

$$EU_{i,1} = \left(p\frac{e_{i,1}+e_{-i,1}}{2} - p^2\frac{e_{i,1}e_{-i,1}}{4}\right)Y - c(e_{i,1}) + \left(1-p\frac{e_{i,1}}{2}\right)\left(1-p\frac{e_{-i,1}}{2}\right)EU_{i,2} \tag{3.4}$$

Compared to joint experimentation, the strategic interaction of the two agents across periods is now determined through a new channel. Agents know they receive a payoff of Y if there is at least one breakthrough. Given that agents work on two separate projects, an agent's incentive to experiment depends on how likely there is a breakthrough in their partner's project, as only one breakthrough is needed. This introduces a component of strategic substitutability between actions within a stage.

Lemma 3. *An agent's second- and her partner's second-stage experimentation efforts as well as an agent's first- and her partner's first-stage experimentation efforts are strategic substitutes.*

Within the second stage, an agent's incentive to experiment decreases in the other agent's experimentation effort, since only one breakthrough is needed to receive Y. There is no benefit in experimenting if the partner achieves a breakthrough, the likelihood of which is increasing in the partner's experimentation effort. The same applies in the first stage.

This strategic substitutability of experimentation within a stage drives the following result concerning experimentation across stages:

[3]Note that the likelihood of a breakthrough in the first stage is given by $p\frac{e_{i,1}}{2} + p\frac{e_{-i,1}}{2} - p\frac{e_{i,1}}{2} \times p\frac{e_{-i,1}}{2} = p\frac{e_{i,1}+e_{-i,1}}{2} - p^2\frac{e_{i,1}e_{-i,1}}{4}$ and correspondingly, the probability of no breakthrough is given by $1-\left(p\frac{e_{i,1}+e_{-i,1}}{2} - p^2\frac{e_{i,1}e_{-i,1}}{4}\right) = \left(1-p\frac{e_{i,1}}{2}\right)\left(1-p\frac{e_{-i,1}}{2}\right)$.

Lemma 4. *An agent's second-stage experimentation increases in their partner's first-stage experimentation.*

As second-stage experimentation of partners are strategic substitutes, an increase in an agent's partner's posterior belief about that partner's project's quality $\rho(e_{-i,1})$ decreases the agent's incentive to experiment, and vice versa. The mechanism behind this is that high experimentation by the partner in the first period will discourage the partner's experimentation in the second period. This is the case, as the partner's posterior about her project's quality decreases in her own first-stage experimentation, which decreases her second-stage experimentation incentives. As within a stage experimentation levels are strategic substitutes, an agent's second-stage experimentation increasing in their partner's first-stage experimentation.

This mechanism operates through a change in the beliefs about the quality of an agent's project associated with changes in that agent's experimentation. An agent's action does not affect the partner's posterior of their own project's quality. Therefore, there exists no informational externality.

Proposition 3. *Experimentation effort is higher in the first stage if this is observable.*

Since an agent's second-stage effort is increasing in their partner's first-stage experimentation, see Lemma 4, and a breakthrough constitutes a public good, the observability of experimentation effort induces higher experimentation levels. The reverse logic from Proposition 1 comes into play here. Now, with unobservable effort the two agents cannot encourage their partner to increase experimentation in the second stage. Therefore, incentives to experiment are higher if this is observable.

Interestingly, Proposition 3 shows that the observability of experimentation effort has the opposite directional effect if partners experiment separately compared to when they experiment jointly (see Proposition 1). A combination of two factors drives this as we move from a setting of joint experimentation to separate experimentation: First, the possibility that experimentation is futile if the partner achieves a breakthrough and second, the lack of an informational externality.

Proposition 4. *Experimentation effort is higher in the second stage if this is unobservable.*

This is again a direct consequence of Proposition 3. In the PBE, beliefs about first-stage experimentation are correct. Agents are now more pessimistic in the second stage if experimentation is observable, because experimentation is higher in the first stage.

The differences in optimal experimentation efforts depending on the observability with separate projects will be small, relative to the effects of observability in joint projects. The observability only matters because agents consider the potential outcome in which both agents achieve a breakthrough in the second stage. This occurs with probability $\rho(e_{i,1})\rho(e_{-i,1})\left(\frac{e_{i,2}e_{-i,2}}{4}\right)$. Considering this potential outcome, the partner's high first-stage experimentation increases incentives to experiment in the second stage, as then the probability of a simultaneous breakthrough decreases. This is because a breakthrough in the partner's project would be less likely, given the lower posterior. In contrast, for joint projects, high first-stage experimentation directly affects the partner's incentives through changes in the posterior beliefs about the project they are experimenting with. Consequently, there are far less stark differences between treatments when experimentation is separate than if experimentation is joint, see Section 3.3.3.

All proofs for this section are presented in Appendix 3.B.

3.3.3 Predictions for experimentation efforts

My experimental parameters are chosen to provide large theoretical treatment differences in the two treatments with joint experimentation, while making sure optimal effort is sufficiently far from 0% and 100% to avoid boundary effects. The theoretically predicted experimentation effort levels for this set of parameters are summarised in the top rows of Table 3.1. For the parameters chosen in the experiment, the efficient experimentation levels are presented in the bottom rows of Table 3.1.[4]

		Observable effort	Unobservable effort
Equilibrium levels	Joint	$e_{i,1} = 10\%$ $e_{i,2} = 77\%$	$e_{i,1} = 34\%$ $e_{i,2} = 65\%$
	Separate	$e_{i,1} = 50\%$ $e_{i,2} = 61\%$	$e_{i,1} = 47\%$ $e_{i,2} = 61\%$
Efficient levels	Joint	$e_{i,1} = 100\%$ $e_{i,2} = 0\%$	$e_{i,1} = 100\%$ $e_{i,2} = 0\%$
	Separate	$e_{i,1} = 74\%$ $e_{i,2} = 100\%$	$e_{i,1} = 74\%$ $e_{i,2} = 100\%$

Notes: The top two rows present experimentation levels in the PBE for the chosen parameters by treatment. The bottom two rows present the efficient experimentation levels for the chosen parameters by treatment.

Table 3.1: Theoretical treatment predictions

[4] If experimentation is joint, joint payoffs are maximized if agents maximize EV:

$$EV = p\left(e_{i,1} + e_{-i,1}\right)Y - c(e_{i,1}) - c(e_{-i,1}) + \left(1 - p\left(\frac{e_{i,1} + e_{-i,1}}{2}\right)\right) \times \left(\rho(e_{i,1}, e_{-i,1})\left(e_{i,2} + e_{-i,2}\right)Y - c(e_{i,2}) - c(e_{-i,2}))\right)$$

If experimentation is separate, joint payoffs are maximized if agents maximize EV:

$$EV = \left(p\left(e_{i,1} + e_{-i,1}\right) - p^2\frac{e_{i,1}e_{-i,1}}{2}\right)Y - c(e_{i,1}) - c(e_{-i,2}) + \left(1 - p\frac{e_{i,1}}{2}\right)\left(1 - p\frac{e_{-i,1}}{2}\right) \times \left(\left(\rho(e_{i,1})e_{i,2} + \rho(e_{-i,1})e_{-i,2} - \rho(e_{i,1})\rho(e_{-i,1})\frac{e_{i,2}e_{-i,2}}{2}\right)Y - c(e_{i,2}) - c(e_{-i,2})\right)$$

3.4 Experimental design

The experimental design closely follows the two theoretical models described in Section 3.3.1 and Section 3.3.2. The study was pre-registered at the AEA RCT Registry (Brütt, 2020). The experiment employs four treatments. I vary in a between-subject 2-by-2 design the observability of experimentation effort and whether experimentation is joint or separate.

In all treatments, subjects repeatedly play an experimentation game. For each of these games, two participants are randomly paired to be in a 'team'. Each team member has to choose how much of their individual budget of €2 to invest in two stages of the experimentation game. They can invest between 0% and 100% of their budget in each stage.[5]

3.4.1 Treatments

The four treatments differ along two dimensions. First, the experiment varies whether experimentation is joint. In the joint case, the two paired subjects work on one project and can achieve a 'breakthrough' depending on the level of joint experimentation. A breakthrough reveals the project's quality and guarantees a payoff for all team members. The incentives for treatments with joint experimentation are as outlined in Section 3.3.1. If experimentation is separate, subjects work on two distinct projects with independently drawn quality. Their individual experimentation determines the likelihood of a breakthrough in their individual project. The incentives for treatments with separate experimentation are discussed in 3.3.2. In all treatments, a breakthrough results in a payoff for both agents.

Second, the treatments differ in the observability of experimentation effort. In treatments with observable experimentation investments, the participants are informed of their team member's investment level after the first stage, before making their own second-stage investment choice. In the treatments without observable experimentation, participants only know how much they invested themselves in the first stage before moving to the second stage.

3.4.2 Experimental timeline

Figure 3.1 illustrates the timing of the experiment. Subjects face 30 rounds of the experimentation game. Each round of the experimentation game starts with the first investment stage. After the first stage, a set of beliefs is elicited, see Section 3.4.4 for details. In treatments with observable experimentation, the participants afterwards

[5]This excludes the possibility of negative payoffs.

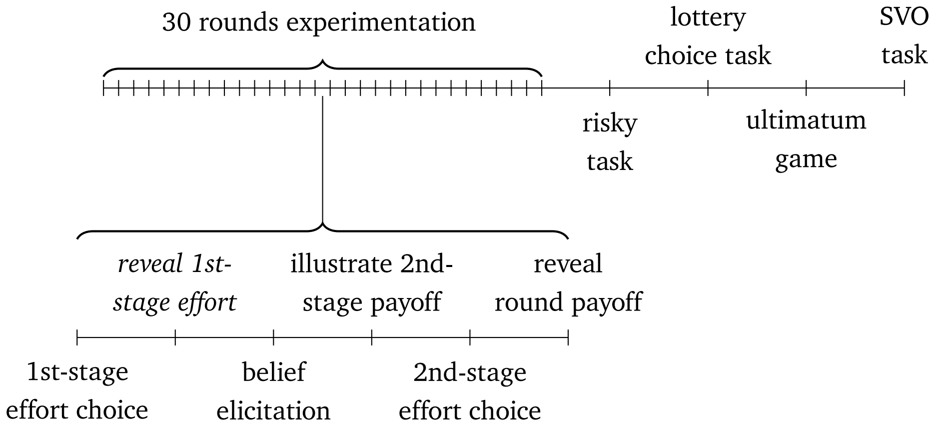


Figure 3.1: Outline of the experiment

learn their partner's first-stage experimentation investment. Furthermore, the participants receive decision support for the second stage, see Section 3.4.3 for details.

Next, subjects make their second-stage investment decision for the case that there was no breakthrough in the first stage, using the strategy method. The strategy method ensures that I collect observations of the second-stage investment even if the project has been terminated due to a breakthrough.[6] At the end of a round, subjects receive feedback. This feedback includes their payoff of the round, whether a breakthrough was achieved, their investment and their partner's investment if this was observable. Afterwards, subjects are re-matched to a new partner, within matching groups of six.

The aim of repeating the game is to facilitate subjects' learning. While the game is rather complex at first, repeating it allows subjects to understand how their behaviour in the first stage can influence beliefs and behaviour in the second stage. After 30 rounds of the experimentation game, subjects go through four separate control tasks that serve to identify potential drivers of experimentation. These tasks are administered after the experimentation game to avoid any impact on the experimentation game itself.

The first control task is a decision under uncertainty. This task is closest to the experimentation game. The subjects face a risky choice to invest in a project. The project can be of high or low quality and the subjects have to choose how much of an endowment to invest. The parameters are the same as in the experimentation game. Being a one-shot game, it excludes learning and informational externalities to focus on the uncertainty about the project's quality in experimentation. There are also no payoff externalities. This task is added as studies such as Banks et al. (1997) do not document any explanatory power of risk aversion measured by standard risk aversion

[6]This in particular guarantees that there are also sufficiently many observations even for high experimentation levels. There would be fewer of these observations otherwise, due the higher likelihood of a breakthrough with high experimentation investment.

elicitations, despite uncertainty being central to experimentation. In contrast, Hudja and Woods (2021) find that risk aversion does entail explanatory power if measured in a setting resembling the actual task more closely. This is in line with the recent findings of Charness et al. (2020), showing that risk preference elicitations remain predictive only in closely related frameworks. I complement this with a standard lottery choice task to obtain a standard measure of the subject's risk attitude (Holt and Laury, 2002). Next, the subjects play an ultimatum game (Güth et al., 1982). The second stage employs the strategy method, where subjects indicate their lowest acceptable offer. Finally, I measure the subjects' social value orientation using the ring test (Liebrand, 1984).

3.4.3 Decision support

The subjects need to understand the consequences of their own actions and their team members' actions on their payoffs. To facilitate this understanding, the subjects are offered a graphical interface that shows the possible payoff consequences of their actions in the second stage. The subjects can enter multiple values of possible second-stage effort levels by their team member and various beliefs they may have about the probability that the project is of high quality. Given these variables, the tool shows the expected payoffs for each possible effort level by the subject. The graph clarifies the consequences of a certain experimentation level for both the own and the other's expected payoff; it does not encourage subjects to choose any specific level.[7]

To ensure that the decision support does not push the subjects to only consider their own payoff, the graph also shows the payoff consequences of choosing a certain experimentation level for the partner. This avoids limiting subjects to maximising the own expected payoffs. If they wish, they can consider other outcome dimensions, such as overall payoffs or inequalities between payoffs, which are equally salient in the graph.

Furthermore, calculators are available in both stages of the experimentation game. These allow the subjects to calculate the costs of investing and the probability of a breakthrough for given investment levels.

Both the graphical interface and the calculators are only shown to participants if they actively choose to reveal them. This way, the subjects can ignore the provided support if they want to. This aims at ensuring that the subjects' true preferences are elicited; payoff consequences are transparent, while the subjects only receive the information they desire. During the instructions, the participants see a video demonstrating how to use the graphical interface and the calculators.

[7] Screenshots of this tool are available in the instructions in Appendix 3.D.

3.4.4 Belief elicitation

The following types of beliefs are elicited after the first stage of the experimentation game:

1. The posterior belief about the project's quality
2. The belief about the partner's posterior belief about the project's quality
3. The belief about the experimentation investment by the partner in the second stage
4. *Only if effort is unobservable:* The belief about the experimentation investment by the partner in the first stage

I use the binarised scoring rule (BSR) introduced in Hossain and Okui (2013) to incentivise the belief elicitations. The chance of receiving a prize of €2 increases in the accuracy of the prediction. For this, a quadratic loss function is used. The BSR ensures that reporting true beliefs is incentive compatible even if the subjects are risk averse or non-expected utility maximisers. This type of scoring rule outperforms non-binarised scoring rules (see e.g. Harrison et al., 2014; Erkal et al., 2020).

Danz et al. (2022) show that using the BSR may give rise to errors in the belief elicitation if the incentivisation is transparent. Therefore, the subjects are only informed that giving their truthful best guess will maximise the probability of receiving the prize for their prediction. Detailed information on the incentivisation is withheld from the subjects, unless requested. See Appendix 3.E for the detailed instructions given to the participants.

3.4.5 Procedures

384 students recruited at the CREED laboratory of the University of Amsterdam participated in this study from September to November 2020. The experiment included 32 sessions, each consisting of 12 subjects in two matching groups per session. The participants did not know the identity of the other participants in their session or matching group. The experiment was advertised as a three-hour experiment on economic decision making, without any further details. The experiment was computerised using PHP. The treatment assignment was randomised evenly at the session level. Upon starting the experiment, the subjects were randomly assigned to matching groups.

The experiment was conducted online due to the COVID restrictions at the time. The participants received a link for the experiment and an invitation to join a zoom session. The zoom session allows the participants to ask the experimenter any questions they may have.[8]

[8] I guaranteed anonymity by re-naming subjects and ensured that no communication was possible

Given that this experiment is online, the participants are more likely to stop the experiment early. If a subject dropped out before the first round of the experiment, I substituted in a back-up player on their behalf.[9] While the experiment was conducted online, the subject pool reflects the standard laboratory population, as the database of enlisted subjects of the CREED laboratory was used for recruitment. It was communicated that practices commonly used at the CREED laboratory, such as a no-deception policy, would also apply online.

The instructions are available in Appendix 3.D. The understanding of these instructions was tested before the start of the experiment. Two rounds of the experimentation game and two other rounds of the belief elicitation were randomly chosen for payment. In addition, all control tasks were paid out. Earnings were on average €32.65. The average duration of the experiment was approximately 2 hours and 21 minutes.

3.5 Experimental results

Table 3.1 and Table 3.2 report the experimentation effort per treatment for the first and the second stage of the experimentation game, respectively. Average experimentation levels are shown by observability (left vs. right column) and by whether experimentation is joint or separate (top vs. bottom row). The results presented here are robust to only considering observations from the second half of the experiment, so not driven by inexperience. This pre-registered robustness check is provided in the Appendix in Tables 3.C.1 and 3.C.2.

		1st stage		
	Unobservable		Observable	
Joint	61.46%	$<^{*}$	70.98%	66.24%
	$\vee^{**}$		$\vee^{***}$	$\vee^{***}$
Separate	50.43%	$<$	56.69%	53.56%
	55.92%	$<^{**}$	63.84%	

Notes: Average experimentation effort in the first stage. Differences in means are tested with Permutation T-tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.1: Experimentation effort per treatment in the first stage

I apply Permutation T-tests (*PmtT-test*) when studying treatment comparisons and comparisons of observed behaviour to the theoretical predictions. These tests are more

between subjects by muting everyone and restricting the chat function to communication only with the experimenter.

[9] This way, the matching groups are not reduced in size and there is no loss in data for the remaining players. As the experience of these back-up players does not differ from the experience of any other participant, I will include their data in the analysis. In total, there were 17 drop-outs *before* the start of the experimentation game for whom back-ups were substituted in. In contrast, I discard the data of participants who dropped out prematurely after the first round of the experimentation game and the data of those that replaced them for the analysis. There was only one drop-out *after* the start of the experimentation game.

	2nd stage			
	Unobservable		Observable	
Joint	33.51%	>	26.74%	30.11%
	$\wedge$		$\wedge^{**}$	$\wedge^{***}$
Separate	38.53%	<	38.88%	38.71%
	36.03%	>	32.81%	

Notes: Average experimentation effort in the second stage. Differences in means are tested with Permutation T-tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.2: Experimentation effort per treatment in the second stage

powerful than traditional non-parametric techniques such as the Wilcoxon signed-rank and Mann-Whitney U tests (Siegel and Castellan, 1981). Given the lack of independence of observations within a matching group, the observations are averaged at the matching-group level. For regression analyses, I cluster the observations at the matching-group level to account for the dependence of observations.

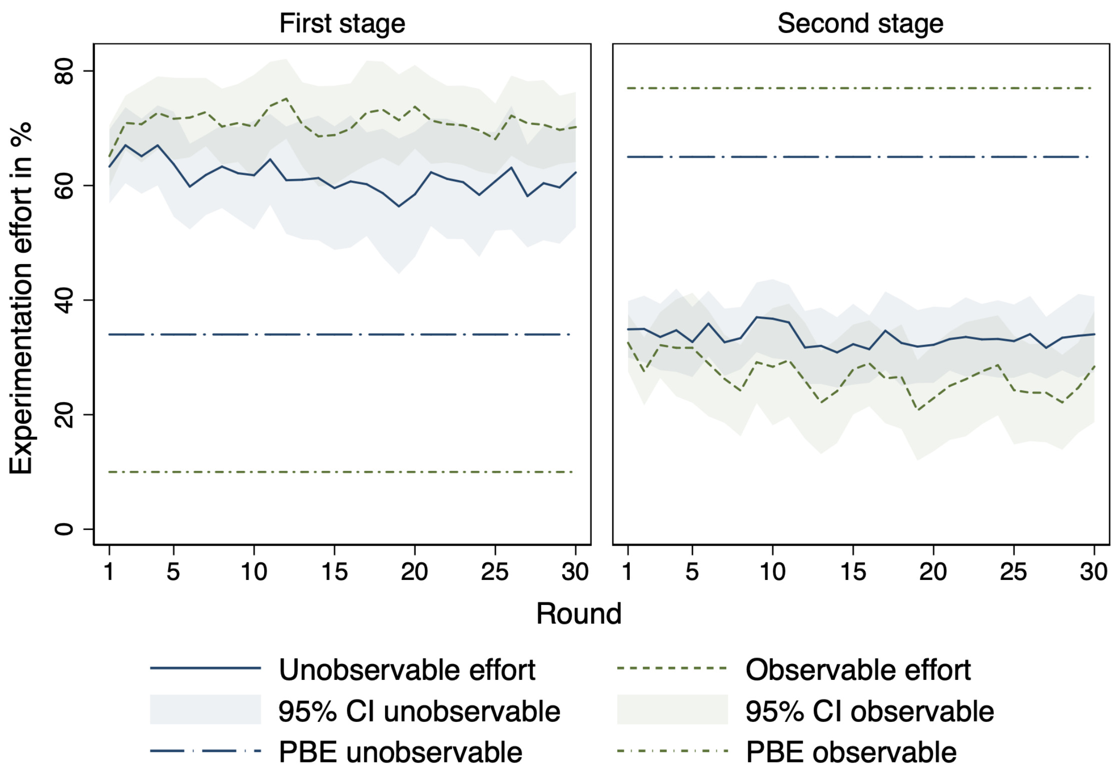
Section 3.5.1 discusses the observed behaviour in the two treatments with joint experimentation in teams, comparing observable and unobservable experimentation. Section 3.5.2 then discusses how this relates to the setting where individuals experiment separately, Section 3.5.3 explores other channels to explain the observed experimentation behaviour, considering joint and separate experimentation together.

3.5.1 The observability of experimentation in joint projects

First, I consider the case of joint experimentation. Here, there are stark difference between the theoretically predicted behaviour and actual behaviour in the laboratory. Figure 3.1 illustrates this for experimentation in the first and second stages of the experimentation game. The left panel sets the first-stage experimentation in both joint treatments against the PBE predictions, the right panel does so for second-stage experimentation.

3.5.1.1 First-stage experimentation in joint projects

As a first step, I consider experimentation in the first stage. It is important to keep in mind that only first-stage experimentation includes some critical elements of experimentation. Here, participants' experimentation effort can generate new information to be used both by themselves and by their partner in the second stage. If effort is observable, participants invest 70.98% in the project, while they invest 61.46% if this is unobservable (see Table 3.1). For both these values, it is evident from Figure 3.1 that I can reject the theoretically predicted experimentation levels of 10% if observable and 34% if unobservable in favour of higher experimentation (*PmtT-test*; both



Notes: Comparison of experimentation in the first stage (left) and second stage (right) over 30 rounds of joint experimentation to the PBE predictions. Shaded regions indicate the 95% confidence interval, clustering observations on a matching group level.

Figure 3.1: Experimentation in treatments with joint experimentation

$p < 0.001$).[10]

Result 1. *First-stage experimentation is higher than predicted in joint experimentation.*

At first glance, this is surprising, given that earlier studies discussing individual experimentation, such as Meyer and Shi (1995), tend to observe under-experimentation. This is also in contrast to Boyce et al. (2016), who study strategic experimentation without a payoff externality. However, there are several distinct features of strategic experimentation with payoff externalities that may help explain this observation, which will be discussed throughout this section.

Strikingly, first-stage experimentation is significantly higher if it is observable, contrary to the theoretical predictions. There is an approximately 15% increase with observable experimentation effort (*PmtT-test*; $p = 0.058$).

Result 2. *First-stage experimentation is higher if experimentation effort is observable.*

[10]To myopically maximise first-stage payoffs, not considering the effect on second-stage payoffs, agents should choose an effort level of $e_{i,1} = \frac{p \times Y}{8} = 81.25\%$. This is significantly higher than the observed experimentation levels for both treatments (*PmtT-test*; both $p < 0.001$). Thus, while agents experiment more than theoretically predicted, there is no evidence that agents are fully myopic either.

Result 2 implies that I can clearly reject the theoretical prediction that the observability of experimentation effort decreases experimentation. This effect is not an artefact of early rounds of experimentation, where subjects are still learning about the exact incentives they face. Instead, in the last half of the experiment experimentation is also 18% higher if observable (*PmtT-test*; $p = 0.046$). Thus, the presence of an informational externality does not decrease experimentation levels.

Several channels can drive higher levels of experimentation than predicted in the first stage if experimentation is observable. To explore why a discouragement effect may not be present, I first focus on the channels that can be identified by studying joint experimentation. First, the lack of a discouragement effect could be explained through biases in belief formation. Second, social preferences, specifically reciprocity, could account for this. These channels are now discussed separately.

3.5.1.2 Belief formation in joint projects

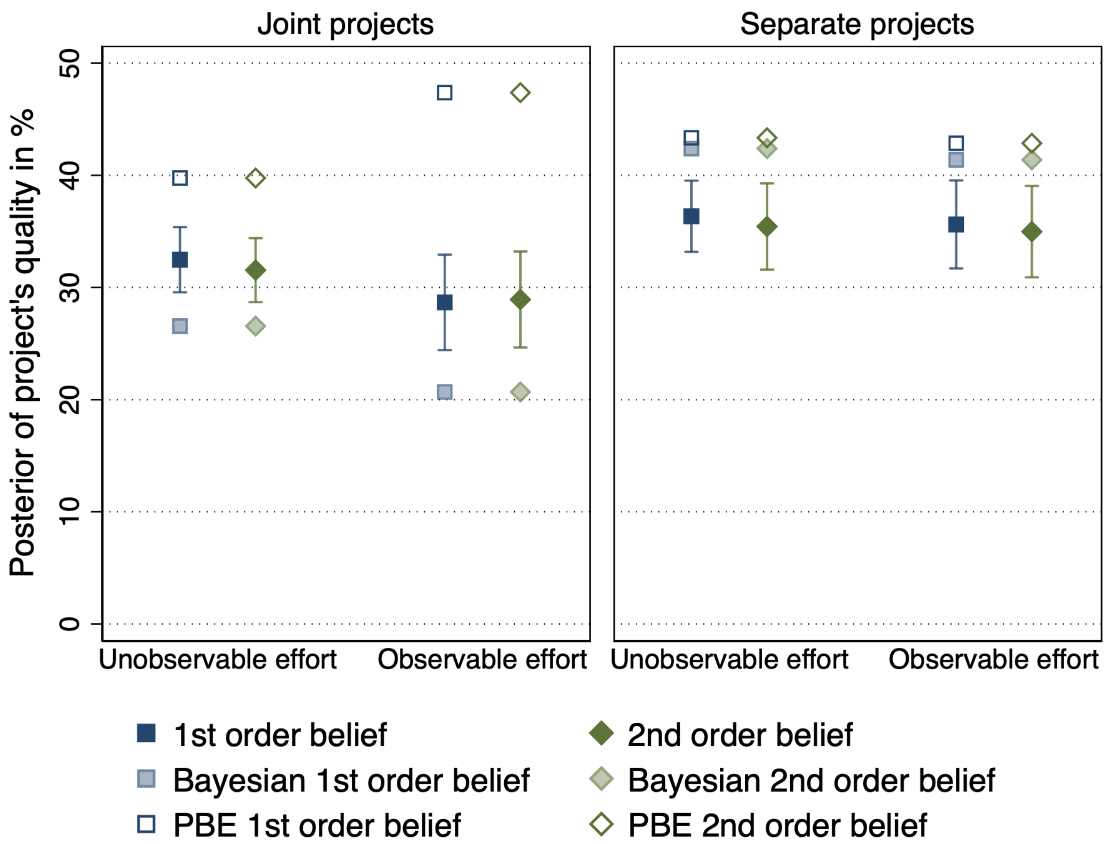
Figure 3.2 contrasts the participants' average beliefs with both the Bayesian posterior given their first-stage experimentation and the beliefs in the PBE for all four treatments. The left panel presents these differences for joint projects, the right for separate projects. Table 3.3 provides the overview of this comparison, contrasting the elicited beliefs both to the beliefs in the PBE and Bayesian posteriors.

	Unobservable		Observable
Joint	32.48% ($>^{***}$*26.56%*,$<^{***}$*39.74%*)	$>$	28.67% ($>^{***}$*20.63%*,$<^{***}$*47.37%*)
	$\wedge^{*}$		$\wedge^{**}$
Separate	36.35% ($<^{***}$*42.38%*,$<^{***}$*43.34%*)	$>$	35.62% ($<^{***}$*41.38%*,$<^{***}$*42.86%*)

Notes: Average posteriors after first-stage experimentation. Treatment differences are tested with Permutation T-tests using matching group averages. In parentheses, the table provides *Bayesian posteriors in italics* and *PBE beliefs in grey italics*. Differences in means and differences to theoretical predictions are tested with Permutation T-tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3: Posterior of the project's quality

There is no significant difference in posteriors between treatments with joint experimentation depending on whether experimentation is observable or not (*PmtT-test*; $p = 0.131$). Remember that since first-stage experimentation is higher if it is observable, agents should become more pessimistic in the observable treatment. However, the experiment may be under-powered to see this reflected in beliefs. In particular, beliefs appear to be updated similarly across treatments. Both in the case of unobservable and of observable effort, beliefs are significantly below the beliefs in the PBE (*PmtT-test*; both $p < 0.001$). This is consistent with higher-than-predicted first-stage experimentation and no evidence for biases in belief updating, as the observed experimentation efforts differ from the PBE predictions.



Notes: Comparison of own (1st order) and beliefs about partner's (2nd order) posteriors in the joint treatments (left) and the separate treatments (right) to PBE beliefs and Bayesian posteriors given the (beliefs about) first-stage experimentation. Bars indicate the 95% confidence interval, clustering observations on a matching group level.

Figure 3.2: Elicited beliefs vs. theoretical predictions

To establish whether biases in belief updating exist, I have to consider the comparison between the elicited beliefs and the Bayesian posteriors. The Bayesian posteriors are calculated based on the empirical first-stage experimentation, if observable, and the subjects' beliefs about their partner's first-stage experimentation, if unobservable. Here, Figure 3.2 illustrates that there is a comparable difference between the elicited beliefs and Bayesian posteriors in both treatments. Beliefs are significantly more optimistic than the Bayesian posterior, both if the first-stage experimentation effort is unobservable and if it is observable (*PmtT-test*; both $p < 0.001$). This suggests that agents update their beliefs conservatively in both treatments, as frequently observed in the literature (see e.g. Benjamin (2019) for a recent overview). Since there is no significant difference in this measure of conservatism between treatments (*PmtT-test*; $p = 0.379$), this does not point towards a lack of first-stage experimentation being able to discourage future experimentation through an (absent) effect on beliefs.

Result 3. *Beliefs are updated conservatively, but respond to experimentation in the predicted direction.*

Treatments	*Dependent variable*: Posterior of project's quality			
	Joint projects		Separate projects	
	Obs	Unobs	Obs	Unobs
	(1)	(2)	(3)	(4)
Own effort	-0.08*	-0.09**	0.06	-0.03
	(0.04)	(0.03)	(0.05)	(0.04)
Partner's effort	-0.16***	0.04	0.07*	0.01
	(0.05)	(0.06)	(0.04)	(0.03)
Constant	46.18***	35.51***	28.14***	37.41***
	(4.74)	(4.09)	(5.45)	(2.75)
Observations	2880	2850	2880	2880
Clusters	16	16	16	16
R-squared	0.055	0.012	0.026	0.002

Notes: OLS estimating effect of own and partner's first-stage experimentation effort on posterior of project's quality for all treatments. (1) and (3) use the partner's actual experimentation effort, (2) and (4) use the subject's belief about the partner's experimentation effort. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: The effect of first-stage experimentation on posterior beliefs

I will now examine more closely how a discouragement effect impacts second-stage experimentation through changes in beliefs. For a discouragement effect to exist in the treatment with observable experimentation, a first necessary condition is that first-stage experimentation affects the subjects' posterior beliefs about the project's quality. More specifically, a participant's posterior belief has to decrease in her partner's first-stage experimentation. Table 3.4 shows the regression results of individuals' beliefs on their own experimentation and their partner's experimentation, when the partner's experimentation is observable, or the elicited belief about the partner's experimentation if unobservable. Own and partner's first-stage experimentation indeed significantly and negatively correlate with the posterior beliefs if experimentation is joint and observable ($p = 0.005$; column (1) in Table 3.4).[11] Observing a one percentage point increase in first-stage experimentation by a subject's partner is associated with a 0.16 percentage point decrease in the posterior about the project's quality.

This is in contrast to the case where experimentation effort is not observable. In that case, there is no correlation between beliefs about the partner's first-stage experimentation and the posterior ($p = 0.543$), see column (2) in Table 3.4. This suggests that the subjects only put weight on the elements they actually observe when forming their beliefs, reflecting the inherent uncertainty about their partner's first-stage experimentation if this is unobservable.

[11]This is explained largely by between-subject variation, not within-subject variation. In Appendix 3.C, I show that subject-level fixed effects absorb the effect of own first-stage experimentation on the posterior beliefs, suggesting that variation in first-stage experimentation and associated changes in beliefs between subject drive the effect of experimentation on beliefs.

Treatments	*Dependent variable:* Belief about partner's posterior of project's quality			
	Joint projects		Separate projects	
	Obs	Unobs	Obs	Unobs
	(1)	(2)	(3)	(4)
Own effort	-0.13***	-0.01	0.05	-0.04
	(0.04)	(0.04)	(0.04)	(0.05)
Partner's effort	-0.13**	0.01	0.13**	0.01
	(0.05)	(0.07)	(0.05)	(0.05)
Constant	47.90***	32.10***	24.84***	36.66***
	(4.50)	(3.84)	(5.13)	(2.80)
Observations	2880	2850	2880	2880
Clusters	16	16	16	16
R-squared	0.057	0.000	0.050	0.003

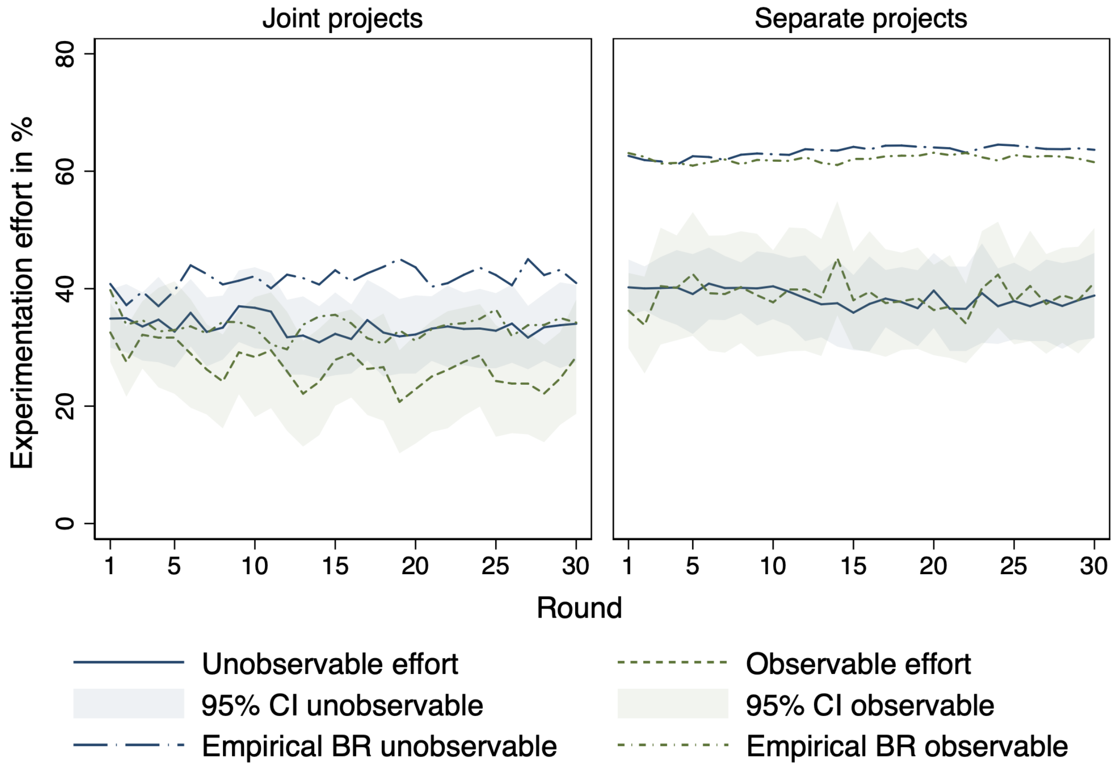
Notes: OLS estimating effect of own and partner's first-stage experimentation effort on beliefs about partner's posterior of project's quality for all treatments. (1) and (3) use the partner's actual experimentation effort, (2) and (4) use the subject's belief about the partner's experimentation effort. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: The effect of first-stage experimentation on beliefs about partner's posterior

A second crucial element of the discouragement effect is that subjects not only update their beliefs in the prescribed manner, but also expect their partners to do so. Only in this case do individuals face an incentive to decrease first-stage experimentation to avoid discouraging future experimentation of their partner. Table 3.5 reports the correlations between on the one hand participants beliefs about their partner's posterior and on the other hand a own experimentation and the partner's experimentation (for the treatment with observable experimentation) or the beliefs about the partner's experimentation (for the treatment with unobservable experimentation). For the case of observable experimentation, column (1) reveals a pattern consistent with individuals correctly anticipating that their own as well as their partners' first-stage experimentation will result in their partner having more pessimistic beliefs. Elicited beliefs are in line with individuals expecting their partner to become 0.13 percentage points more pessimistic if their first-stage experimentation increases by one percentage point. This is a statistically significant correlation ($p = 0.002$). There is no such correlation if experimentation is observable ($p = 0.680$), see column (2) in Table 3.5 . This indicates that the participants anticipate the potential of discouraging their partner if they choose high experimentation levels if experimentation is observable.

3.5.1.3 Second-stage experimentation in joint projects

Next, consider second-stage experimentation. Here, there should theoretically be a response to the experimentation from the first stage. In particular, the informational



Notes: Comparison of experimentation in the second stage in treatments with joint experimentation (left) and separate experimentation (right) over 30 rounds of experimentation to the empirical best response. Shaded regions indicate the 95% confidence interval, clustering observations on a matching group level.

Figure 3.3: Second stage experimentation vs. empirical best responses

spillovers should affect behaviour if experimentation is observable. Figure 3.1 illustrates that experimentation is lower in the second stage if this is observable, albeit insignificantly so (*PmtT-test*; $p = 0.116$). Compared to the theoretical predictions, Figure 3.1 shows that effort is 48% lower than theoretically predicted if unobservable and 65% lower if observable (*PmtT-test*; both $p < 0.001$). Result 4 summarises this.

Result 4. *Second-stage experimentation is significantly lower than theoretically predicted.*

The lower second-stage experimentation in the treatment with observable experimentation is a consistent response to high first-stage experimentation. Given first-stage experimentation and the resulting Bayesian posterior, I can calculate the empirical best response for each individual in each round of experimentation. Figure 3.3 plots the empirical best response for each treatment. The left panel compares the empirical best response to actual second-stage experimentation for the treatments with joint experimentation. Second-stage experimentation is significantly lower than the empirical best response (*PmtT-test*; $p = 0.014$). The degree of deviation from the best response is indistinguishable between the two treatments (*PmtT-test*; $p = 0.793$). The

Treatments	*Dependent variable:* Second-stage experimentation effort							
	Joint projects		Separate projects		Joint projects		Separate projects	
	Obs	Unobs	Obs	Unobs	Obs	Unobs	Obs	Unobs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Posterior	0.66***	0.54***	0.64***	0.39***	0.42***	0.22***	0.32***	0.18**
	(0.04)	(0.11)	(0.13)	(0.10)	(0.05)	(0.03)	(0.09)	(0.06)
Constant	7.80***	15.85***	16.22**	24.50***	14.58***	26.48***	27.32***	31.98***
	(2.56)	(3.23)	(5.79)	(2.88)	(1.56)	(1.11)	(3.10)	(2.29)
Fixed effects	✗	✗	✗	✗	✓	✓	✓	✓
Observations	2880	2850	2880	2880	2880	2850	2880	2880
Clusters	16	16	16	16	16	16	16	16
R-squared	0.232	0.138	0.126	0.056	0.510	0.710	0.618	0.660

Notes: OLS estimating effect of posterior about project's quality on second-stage experimentation effort for all treatments. (1)-(4) do not include subject fixed effects, (5)-(8) include subject fixed effects. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: The effect of posterior beliefs on second-stage experimentation

low second-stage experimentation is a first indicator that agents respond to information previously generated, and potentially to informational spillovers.

For the discouragement effect to induce lower experimentation through a change in beliefs, second-stage experimentation must be responsive to a change in posteriors when experimentation is observable. If this were not the case, agents would have no reason to fear discouragement when deciding on first-stage experimentation, knowing that the potential pessimism of their partner does not manifest itself in different actions. Table 3.6 gives the results of regressing second-stage experimentation on the posteriors of the project's quality. Second-stage experimentation responds significantly to the posterior in the predicted direction in all treatments. In particular, individuals invest 0.66 percentage points less experimentation effort if they are one percentage point more pessimistic when effort is observable and agents experiment jointly ($p < 0.001$), see column (1) in Table 3.6.

Interestingly, I can exclude that this effect is entirely driven by variation at the subject level. Instead, there is within-subject variation in beliefs across rounds that affects second-stage experimentation effort. To see this, consider the case where subject fixed effects are included in the regression, see column (5) in Table 3.6. Including subject fixed effects controls for different levels of experimentation and beliefs across subjects, which implies that the remaining effect (0.42 percentage points) on second-stage experimentation stems from variation in a subjects' beliefs across rounds.

The final element required for a change in beliefs to yield a discouragement effect is that the agents also anticipate that their partners respond to changes in their posterior. Individuals might not decrease experimentation levels as a response to the

Treatments	*Dependent variable:* Beliefs about partner's second-stage experimentation effort							
	Joint projects		Separate projects		Joint projects		Separate projects	
	Obs	Unobs	Obs	Unobs	Obs	Unobs	Obs	Unobs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Posterior	0.70***	0.71***	0.70***	0.37***	0.58***	0.44***	0.46***	0.26***
	(0.04)	(0.07)	(0.10)	(0.08)	(0.05)	(0.06)	(0.10)	(0.06)
Constant	7.21**	10.22***	14.60***	22.17***	10.51***	18.67***	22.97***	26.16***
	(2.47)	(3.00)	(4.43)	(2.60)	(1.47)	(1.75)	(3.47)	(2.26)
Fixed effects	✗	✗	✗	✗	✓	✓	✓	✓
Observations	2880	2850	2880	2880	2880	2850	2880	2880
Clusters	16	16	16	16	16	16	16	16
R-squared	0.309	0.268	0.191	0.076	0.527	0.685	0.504	0.618

Notes: OLS estimating effect of belief about partner's posterior about project's quality on second-stage experimentation effort by partner for all treatments. (1)-(4) do not include subject fixed effects, (5)-(8) include subject fixed effects. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: The effect of beliefs about partner's posterior on beliefs about partner's second-stage experimentation

observability of experimentation because they fail to realise that their partner's induced pessimism will make her or him experiment less in the second period.

We can measure this by looking at the correlation between a subject's beliefs about their partner's posterior and their beliefs about the partner's second-stage experimentation. Table 3.7 reports such regression results for all treatments. Evidently, the belief measures support that subjects expect their partners to respond to their posteriors, with an estimated reduction of 0.7 percentage points in the beliefs about the partner's second-stage experimentation resulting from a one percentage point change in the beliefs about the partner's posterior ($p < 0.001$; column (1) in Table 3.7). The strength of the correlation is thus similar to that of the subject's own beliefs and second-stage experimentation. The fixed-effect regression in column (5) again reveals that there is a substantial within-subject response.

All necessary factors that are required for the discouragement effect to decrease experimentation incentives are thus present. This gives Result 5.

Result 5. *If experimentation is observable, high first-stage experimentation discourages high second-stage experimentation through a change in beliefs.*

We can therefore conclude that Result 2 (that establishes that observability increases first-round experimentation) is not a consequence of a lack of sophistication in belief updating or a lack of strategic sophistication in neglecting the impact of own first-stage experimentation on the partner's future behaviour. For an alternative explanation, I now turn to whether conditional cooperation is better suited to account for the observed experimentation patterns.

3.5.1.4 Reciprocal behaviour

Social preferences, specifically reciprocal behaviour, could provide an explanation for higher experimentation levels if these are observable. Given the positive payoff externality, subjects may reward observing high first-stage experimentation with high second-stage experimentation. To test this, I will employ a two-step procedure. This will allow me to see whether high first-stage experimentation is indeed followed by the partner choosing higher second-stage experimentation, separating the direct effect of first-stage on second-stage experimentation from the indirect effect through a change in beliefs.

In the first step, I regress second-stage experimentation $E^2_{i,t}$ in round t of individual i on posterior beliefs $\rho_{i,t}$, clustering standard errors at the matching group level.

$$E^2_{i,t} = \beta\rho_{i,t} + \epsilon_{i,t}$$

In the second step, the residuals of the first regression $\hat{\epsilon}_{i,t}$ are regressed on i's partner's ($-i$) first-stage experimentation $E^1_{-i,t}$ in round t, $Observable_i$, indicating whether i is in the observable treatment, and the interaction of these two variables. When experimentation is unobservable, $E^1_{-i,t}$ is given by i's belief of her partner's first-step experimentation.

$$\hat{\epsilon}_{i,t} = \gamma_1 E_{-i,t} + \gamma_2 Observable_i + \gamma_3 E^1_{-i,t} \times Observable_i + u_{i,t}$$

This ensures that I only capture the direct effect of an individuals first-stage experimentation on their partner's second stage experimentation, which is unrelated to how beliefs are affected. Table 3.8, column (1) provides the results of this second-stage estimation.

Subjects' second-stage experimentation responds significantly to (beliefs of) their partner's first-stage experimentation ($p = 0.048$). However, this effect is entirely driven by the treatment where experimentation is unobservable. The negative interaction effect of the same approximate size ($p = 0.022$) implies that in the treatment with observable experimentation, there is no correlation between the partner's first-stage experimentation and their second-stage experimentation beyond the effect driven by a change in posteriors about the project's quality.

The correlation of first- and second-stage experimentation if experimentation is unobservable, controlling for belief effects, does not suggest a reciprocal motive. More likely, subjects with high experimentation levels expect others to experiment more as well. This is consistent with the fact that there is also a positive correlation between their own first-stage experimentation and their second-stage experimentation

	Dependent variable: Residuals of 1st-stage regression	
	(1)	(2)
Observable	13.94**	8.46
	(6.78)	(6.33)
Partner's effort	0.21**	
	(0.09)	
Observable × Partner's effort	-0.24**	
	(0.10)	
Own effort		0.07
		(0.07)
Observable × Own effort		-0.13
		(0.08)
Constant	-11.97**	-4.07
	(5.57)	(5.53)
Observations	5730	5730
Clusters	32	32
R-squared	0.020	0.007

Notes: OLS estimating difference-in-difference in the effect of first-stage experimentation between treatments on residuals of regression of second-stage experimentation on posteriors. In (1), partner's first-stage and own second-stage experimentation is used. In (2), own first stage and beliefs about partner's second stage experimentation is used. Elicited beliefs are used for observations from unobservable treatment. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8: The effect of (beliefs about) first-stage on second-stage experimentation

($p = 0.017$), also controlling for belief effects using the two-step estimation.[12] Thus, since there is no effect in the observable treatment, there is no evidence of subjects punishing or rewarding their partner's experimentation by increasing their own experimentation.

While there is no reciprocal behaviour, it is conceivable that subjects still expect their partners to reciprocate high experimentation and thus face an incentive to increase first-stage experimentation. If this is the case, beliefs about the partner's second-stage experimentation should increase in own first-stage experimentation. The same two-step procedure is employed as there again exists a belief channel through which first-stage experimentation can affect beliefs. The results of the second-stage regression are presented in Table 3.8, column (2). No significant correlation between own first-stage experimentation and the beliefs about the partner's second-stage experimentation exists ($p = 0.364$).

In line with the preceding analysis, I show in Appendix 3.C Table 3.C.6 that there is no differential correlation between the elicited measure of negative reciprocity from the ultimatum game and second-stage experimentation depending on whether experi-

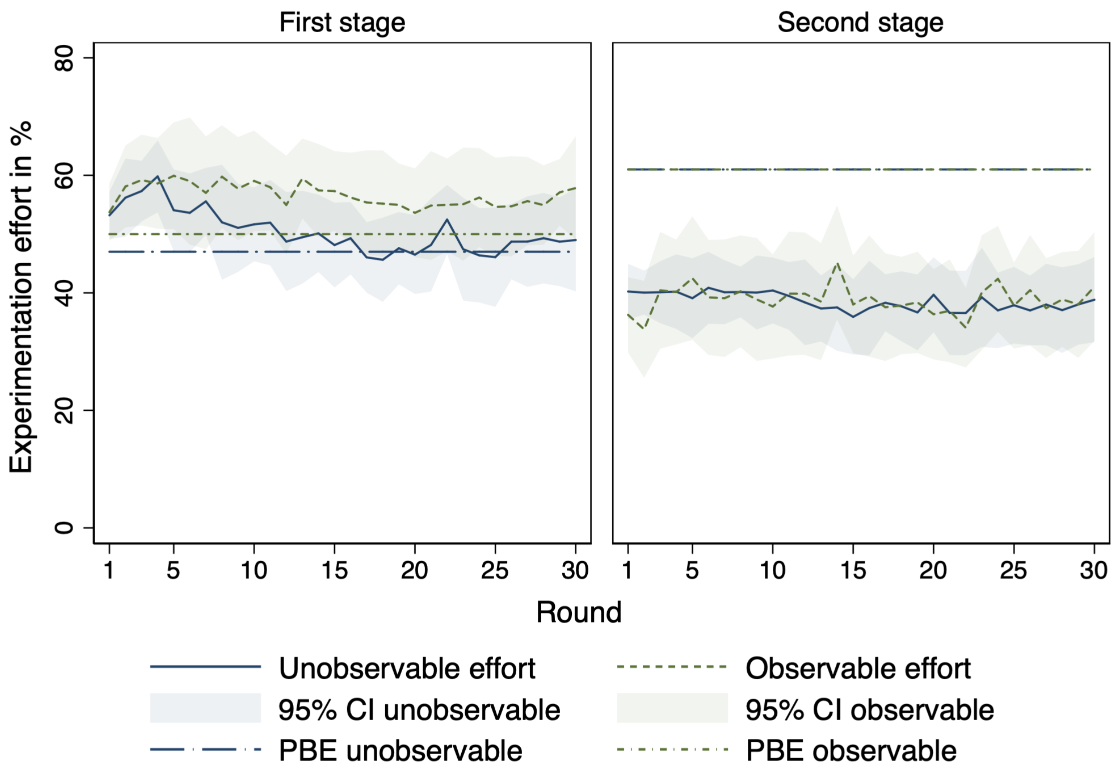
[12] See Appendix 3.C, Table 3.C.5 for the regression results.

mentation is observable or not. There is also no significant difference in the correlation between proposer behaviour in the ultimatum game and first-stage experimentation by treatment. Reciprocity, therefore, does not seem to be driving the fact that first-stage experimentation is higher if effort is observable.

Result 6. *There is no evidence that (expected) reciprocity drives first-stage experimentation if this is observable.*

3.5.2 Separate experimentation compared to joint experimentation

A comparison between joint and separate experimentation may shed further light on how the determinants of experimentation, contrasting settings with and without informational externalities. Figure 3.4 displays first- and second-stage experimentation compared to the PBE experimentation levels in the two treatments where individuals work on separate projects. In Table 3.9, first-stage experimentation levels are regressed on treatment indicators for observable and joint experimentation and their interaction, Table 3.10 reports this for second-stage experimentation.



Notes: Comparison of experimentation in the first stage (left) and second stage (right) over 30 rounds of separate experimentation to the PBE predictions. Shaded regions indicate the 95% confidence interval, clustering observations on a matching group level.

Figure 3.4: Experimentation in treatments with separate experimentation

First, both for observable and for unobservable experimentation, experimentation is significantly lower in the first stage if agents work on separate projects compared to joint experimentation, see column (2) in Table 3.9. For unobservable experimentation, experimentation is 22% higher with joint than with separate experimentation, for observable experimentation 25% higher (*PmtT-test*; $p = 0.028$ and $p = 0.002$, respectively). Joint experimentation clearly has a positive effect on experimentation levels. This gives Result 7.

Result 7. *First-stage experimentation is higher if agents experiment jointly.*

This is in contrast to the theoretical predictions, as the lack of an informational externality implies that significantly higher experimentation levels are expected with separate experimentation if experimentation is observable. Instead, this finding is in line with comparative statics predictions that follow from agents aiming for efficient experimentation levels, maximizing their joint payoffs. Here, separate experimentation leads to lower experimentation in the first stage, as high first-stage experimentation increases the probability of two breakthroughs, which is inefficient. With joint experimentation, two breakthroughs are not possible. Investing fully in the first stage is efficient, thereby resolving all uncertainty.

To test whether the observability of experimentation effort has a distinct effect depending on whether experimentation is joint, consider the interactions in Table 3.9 and Table 3.10. Both for the first and for the second stage, there is no statistically significant differential effect of experimentation observability on experimentation effort depending on whether experimentation is joint or separate ($p = 0.606$ in column (3), Table 3.9 and $p = 0.236$ in column (3), Table 3.10, respectively). This is against the theoretical predictions; the observability of experimentation effort is predicted to increase experimentation if separate, but decrease it if experimentation is joint. Instead, the observability of experimentation overall increases experimentation levels in the first stage ($p = 0.029$), see column (1) in Table 3.9. Hence, the presence of an informational externality does not have have a differential impact on experimentation efforts if this information is observable or not. Thus, an environment of observable experimentation encourages experimentation, irrespective of whether the group members work on a joint or separate projects.

Result 8. *Observable experimentation increases first-stage experimentation, independent of whether experimentation is joint or separate.*

Recall that for separate experimentation, marginally higher first-stage experimentation is expected if this is observable, because this can encourage future experimentation. Considering second-stage experimentation if this is separate, Figure 3.3 shows that experimentation is clearly lower than the empirical best response, both for observable and unobservable experimentation (*PmtT-test*; both $p < 0.001$). This is inconsistent with an encouraging force of higher first-stage experimentation. In line with

	Dependent variable: First-stage experimentation					
	(1)	(2)	(3)	(4)	(5)	(6)
Observable	7.92**		6.26	7.97**		6.95*
	(3.53)		(4.14)	(3.44)		(3.85)
Joint		12.68***	11.02**		12.26***	11.23**
		(3.30)	(4.68)		(3.28)	(4.57)
Joint × Observable			3.26			2.06
			(6.30)			(6.04)
Constant	55.92***	53.56***	50.43***	33.31***	31.12***	26.92***
	(2.53)	(2.14)	(2.60)	(8.87)	(8.13)	(8.34)
Controls	No	No	No	Yes	Yes	Yes
Observations	11490	11490	11490	11010	11010	11010
Clusters	64	64	64	64	64	64
R-squared	0.018	0.045	0.064	0.063	0.087	0.105

Notes: OLS estimating effect of joint experimentation, observability of experimentation and the interaction on first-stage experimentation. (1)-(3) do not include controls variables for individual characteristics, (4)-(6) do. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: First-stage experimentation in all treatments

this, the deviation from the best response is significantly larger when experimentation is joint (*PmtT-test*; $p < 0.001$). Experimenting jointly has a positive effect on experimentation levels in both stages of the game, but this effect is not larger when it is predicted to be.

Compared to the treatments with joint experimentation, there is a stark contrast in how beliefs are updated in projects with separate experimentation. With separate experimentation, individuals' posterior beliefs do not respond to their own first-stage experimentation ($p = 0.309$ if experimentation is observable and $p = 0.465$ if experimentation is unobservable, column (3) and (4), Table 3.4). As shown in Figure 3.2, these beliefs are more pessimistic than the Bayesian benchmark, independent of whether experimentation is observable (*PmtT-test*; $p = 0.002$ if observable and $p < 0.001$ if unobservable). The participant's beliefs are consistent with higher experimentation levels than the ones observed. Given these pessimistic beliefs, the low second-stage experimentation levels in both treatments with separate experimentation are not surprising, significantly below the best response that is based on the Bayesian posteriors (*PmtT-test*; both $p < 0.001$), see Figure 3.3.

3.5.3 Norms of high experimentation and leading by example

An intuitive explanation for higher experimentation levels with joint, observable experimentation is that both these aspects foster a stronger sense of group membership and allow teams to establish norms of high experimentation. This would be compa-

	Dependent variable: Second-stage experimentation					
	(1)	(2)	(3)	(4)	(5)	(6)
Observable	-3.22		0.35	-2.13		2.08
	(3.20)		(4.41)	(3.17)		(4.34)
Joint		-8.60***	-5.03		-8.62***	-4.36
		(3.04)	(3.34)		(3.04)	(3.26)
Joint × Observable			-7.12			-8.41
			(5.95)			(5.83)
Constant	36.03***	38.71***	38.53***	13.58*	17.15**	16.37**
	(1.73)	(2.20)	(2.37)	(7.39)	(7.39)	(7.28)
Controls	No	No	No	Yes	Yes	Yes
Observations	11490	11490	11490	11010	11010	11010
Clusters	64	64	64	64	64	64
R-squared	0.003	0.020	0.026	0.021	0.039	0.045

Notes: OLS estimating effect of joint experimentation, observability of experimentation and the interaction on second-stage experimentation. (1)-(3) do not include controls variables for individual characteristics, (4)-(6) do. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10: Second round experimentation in all treatments

rable to the observability of individual contributions to a public good increasing such contributions (see e.g. Andreoni and Petrie, 2004), even without direct punishment. While this experiment does not include direct elicitations of norms, two pieces of evidence support this argument. First, the variance in first-stage experimentation is lower if experimentation levels are observable (*PmtT-test*; $p = 0.030$). Measured as the variance of decisions within a matching group, the lower variance in matching groups that are exposed to observable experimentation suggests that these groups coordinate on effort levels.[13]

Second, agents adapt their experimentation efforts to previously observed experimentation. While Section 3.5.1 demonstrates that agents do not reciprocate across the two stages of the experimentation game, individuals exert higher experimentation efforts if they have observed high experimentation in earlier rounds ($p = 0.059$; column (2) in Table 3.C.7). Furthermore, if experimentation is not observable, the participants' beliefs about their team member's experimentation are (correctly) below the experimentation levels that participants in the treatments with observable experimentation experience (*PmtT-test*; $p = 0.002$)[14]. This hampers the successful coordination on high experimentation levels.

[13]Interestingly, there is not only a lower variance in groups with observable experimentation in late periods of the game, when groups have frequently observed each others' experimentation, but also in the first half of the experiment (*PmtT-test*; $p = 0.009$ in the first 15 periods). Thus, it does not seem necessary for coordination that agents see experimentation efforts frequently.

[14]Beliefs about the partner's first-stage experimentation do not significantly differ from the actual experimentation levels (*PmtT-test*; $p = 0.346$).

The positive response to the partner's experimentation incentivises agents to 'lead by example'. As in the leading-by-example literature, high experimentation efforts, if observable, can induce high experimentation levels in future periods through two channels. First, leading by example has a signalling value (Potters et al., 2007). High experimentation can signal the (private) belief that investment in the project is lucrative. While there is no asymmetric information, communicating private beliefs about whether experimentation is fruitful can be informative in a complex setting when the participants are unsure of their optimal actions. In line with this, in the setting of separate experimentation, individuals update more positively about the project's quality if they observe high experimentation efforts by their partner ($p = 0.055$; column (3) in Table 3.4). As the projects are independent, the partner's experimentation does not reveal any information about the project's objective quality. However, it could signal that the partner believes the project is a worthy investment, a helpful signal in a complex environment. This only works with observable experimentation and can, therefore, potentially explain why experimentation levels are higher in this case. Second, leading by example can result in reciprocal behaviour in later rounds (Meidinger and Villeval, 2002). As discussed, there is not observable reciprocal behaviour within one round of experimentation, but there is evidence of agents responding to earlier experimentation by their partners in later rounds, in particular, if this is observable.

3.6 Conclusion

This paper studies the two-dimensional free-riding problems inherent in strategic experimentation of teams, examining the type of environments that foster successful experimentation. I consider two dimensions of the experimentation environment: the observability of experimentation, and whether agents work on one joint project or on two separate projects. The observability of experimentation efforts is predicted to decrease experimentation levels when agents experiment with a joint project; this is driven by the presence of an informational externality. Agents are predicted to discourage each other from experimenting if they observe each others' experimentation levels but do not observe a breakthrough. With two separate projects, however, the predictions flip, as there is no informational externality, and the potential of a breakthrough in another project implies that agents want to avoid futile experimentation that results in two breakthroughs.

This study employs an experiment to test these theoretical predictions and identify behavioural drivers of experimentation in teams. Strikingly, teams are capable of largely overcoming the free-riding problem that lies at the core of strategic experimentation. In contrast to the prevalent finding in laboratory experiments that experimentation is undervalued, I find that teams experiment more than predicted.

Though their level of strategic sophistication allows individuals to grasp the discouraging effect of their experimentation with joint projects, experimentation is higher if it is observable and agents experiment more with a joint project. This is not a result of agents punishing or rewarding certain experimentation behaviour. Instead, agents can coordinate on higher effort levels if experimentation is observable. The findings are in line with agents choosing to lead by example if their team member can observe their experimentation. Moreover, the higher experimentation with joint projects suggests that agents aim for not purely individually-optimal experimentation, but instead consider efficient experimentation levels. With joint experimentation, a full resolution of uncertainty is possible and efficient in the first stage, while the possibility of having two breakthroughs with separate projects implies lower efficient experimentation levels.

To conclude, there are mechanisms in place that help teams overcome the theoretical hurdles to experimentation. Teams are able to innovate even in settings where it is in every team members' material interest to decrease their experimentation, as this will discourage others from experimenting in the future. For the bigger picture, we can be cautiously optimistic that having teams active in innovative processes (as opposed to individuals) will not create excessive free-riding and a lack of information discovery, but instead might induce team members to work harder for their fellow team members, giving rise to more innovation. Instead of discouraging team members, informational externalities may even signal high hopes for the project's success, encouraging high experimentation.

APPENDIX

3.A Proofs for Section 3.3.1

The results of this Section apply for parameter regions with internal solutions in both stages of the experimentation game, as used in the experiment. For all proofs, I will consider the case of observable experimentation (Obs) unless otherwise noted.

Lemma 1. *An agent's first- and second-stage experimentation efforts are strategic substitutes.*

To see that an agent's first and second stage experimentation are strategic substitutes for large enough Y, take the cross derivative of the expected utility $\frac{\partial^2 EU_{i,1}}{\partial e_{i,2} \partial e_{i,1}}$:

$$\frac{\partial^2 EU_{i,1}}{\partial e_{i,2} \partial e_{i,1}} = \frac{1}{4}\left(2pc'(e_{i,2}) - pY\right)$$

Further note that second-order partial derivatives are continuous, as $\frac{\partial^2 EU_{i,1}}{\partial e_{i,2}^2} = 2p(e_{i,1} + e_{-i,1}) - c''(e_{i,2})$ and $\frac{\partial^2 EU_{i,1}}{\partial e_{i,1}^2} = -c''(e_{i,1})$ and $c(e_{i,1})$ is twice continuously differentiable. By Young's theorem, there is therefore equality of mixed partials here. Thus, this gives

$$\frac{\partial^2 EU_{i,1}}{\partial e_{i,1} \partial e_{i,2}} = \frac{1}{4}\left(2pc'(e_{i,2}) - pY\right)$$

For strategic substitutes between $e_{i,1}$ and $e_{i,2}$, $\frac{1}{4}\left(2pc'(e_{i,2}) - pY\right) < 0$ has to hold. For this to hold, it is a sufficient condition that $c'(e_{i,2}) < \frac{Y}{2}$, which was assumed. □

Lemma 2. *An agent's and her partner's first- and second-stage experimentation efforts are strategic substitutes.*

To see that an agent's first and and her partner's second stage experimentation are strategic substitutes, take the cross derivative $\frac{\partial^2 EU_{i,1}}{\partial e_{-i,2}\partial e_{i,1}}$

$$\frac{\partial^2 EU_{i,1}}{\partial e_{-i,2}\partial e_{i,1}} = -\frac{pY}{4}$$

Again, this is also the cross derivative $\frac{\partial^2 EU_{i,1}}{\partial e_{i,1}\partial e_{-i,2}}$:

$$\frac{\partial^2 EU_{i,1}}{\partial e_{i,1}\partial e_{-i,2}} = -\frac{pY}{4}$$

As $-\frac{pY}{4} < 0$, I can conclude that an agent's first and and her partner's second stage experimentation are strategic substitutes. □

Proposition 1. *Experimentation effort is higher in the first stage if this is unobservable.*

Consider the case where effort is observable. In the PBE, an agent will choose first-stage experimentation such that the marginal benefits from experimentation equal the marginal costs from experimentation, given the other player's action and their beliefs. Part of the costs of increasing experimentation are that the partner will decrease second-stage experimentation, because of the strategic substitutability.

Formally, this means that first-stage experimentation is chosen according to the following first order condition:

$$\frac{\partial EU_{i,1}}{\partial e_{i,1}} = \frac{pY}{2} - c'(e_{i,1}) + \left(1 - p\left(\frac{e_{i,1}+e_{-i,1}}{2}\right)\right)\frac{\partial EU_{i,2}}{\partial e_{i,1}} - \frac{p}{2}EU_{i,2} = 0$$

Next consider the case where effort is not observable. If effort is unobservable, this implies that agents base their decisions on their beliefs about their partner's first-stage effort $\hat{e}_{-i,1}$ instead of the actual partner's effort $e_{-i,1}$ when deciding on optimal effort in the second stage. Simultaneously, their partner will base their decisions on their beliefs about their partner's first-stage effort $\hat{e}_{i,1}$ instead of the actual effort $e_{i,1}$. In the above expression, this may affect experimentation through changes in the terms $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$ and $EU_{i,2}$.

Assume agents do not observe their partner's experimentation but exert experimentation efforts that correspond to the effort levels in the PBE with observable effort. In this case, there is a profitable deviation to exert more effort in the first stage. To see this, consider how the two terms $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$ and $EU_{i,2}$ compare for the two cases.

In the observable case:

$$\frac{\partial EU_{i,2}}{\partial e_{i,1}} = \frac{\partial \rho(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} \left(\frac{e_{i,2} + e_{-i,2}}{2} \right) Y + \rho(e_{i,1}, e_{-i,1}) \frac{\partial e_{i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} \frac{Y}{2} +$$
$$\rho(e_{i,1}, e_{-i,1}) \frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} \frac{Y}{2} - c'(e_{i,2}) \frac{\partial e_{i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}}$$

Now, I consider how this term changes, while holding constant that agents exert experimentation efforts that correspond to the effort levels in the PBE. In the unobservable case, $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$, $\rho(e_{i,1}, e_{-i,1}) \frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} \frac{Y}{2}$ is zero, as the unobservability of $e_{i,1}$ implies $\frac{\partial e_{-i,2}(\hat{e}_{i,1}, e_{-i,1})}{\partial e_{i,1}} = 0$. As $\rho(e_{i,1}, e_{-i,1}) \frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} \frac{Y}{2} < 0$ when effort is observable, see Lemma 2, we have $\frac{\partial EU_{i,2}}{\partial e_{i,1}}^{Obs} < \frac{\partial EU_{i,2}}{\partial e_{i,1}}^{Unobs}$ for a given $e_{i,1}$. Therefore, effort observability decreases incentives to experiment through a change in $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$.

In the case that agents exert experimentation efforts that correspond to the effort levels in the PBE with observable experimentation, $EU_{i,2}$ is the same between the cases with observable and unobservable effort, as posterior beliefs $\rho(e_{i,1}, e_{-i,1})$ will be the same. Thus, taken together, there is a profitable deviation to experiment more with unobservable experimentation efforts.

Hence, if there is an interior solution, first-stage experimentation efforts in the PBE with unobservable experimentation ($e_{i,1}^{Unobs}$) are higher than in the PBE with observable experimentation ($e_{i,1}^{Obs}$). □

Proposition 2. *Experimentation effort is higher in the second stage if first-stage experimentation effort is observable.*

The PBE requires, by sequential rationality, that the agent will maximise her expected utility given her beliefs. Thus, in the case where effort is observable, an agent will choose second-stage effort according to the following condition:

$$\frac{\rho(e_{i,1}, e_{-i,1})}{2} Y = c'(e_{i,2}) \tag{3.5}$$

If effort is unobservable, the agent will choose second-stage effort according to the following condition:

$$\frac{\rho(e_{i,1}, \hat{e}_{-i,1})}{2} Y = c'(e_{i,2}) \tag{3.6}$$

In the PBE, agents have correct beliefs about their partner's first-stage experimentation, implying that $\hat{e}_{-i,1} = e_{-i,1}$. Given that both $e_{i,1}$ and $e_{-i,1}$ are higher with unobservable than with observable effort, see Proposition 1, this gives

$$\rho\left(e_{i,1}^{Unobs}, \hat{e}_{-i,1}^{Unobs}\right) < \rho\left(e_{i,1}^{Obs}, e_{-i,1}^{Obs}\right)$$

As $c''(e_{i,2}) > 0$, $e_{i,2}^{Unobs} < e_{i,2}^{Obs}$ has to hold such that Eqs. 3.5 and 3.6 are both satisfied. □

3.B Proofs for Section 3.3.2

The results in this Section apply for the parameters chosen in the experiment. For all proofs, I will consider the case of observable experimentation (Obs) unless otherwise noted.

Lemma 3. *An agent's second- and her partner's second-stage experimentation efforts as well as an agent's first- and her partner's first-stage experimentation efforts are strategic substitutes.*

To see that an agent's second and her partner's second-stage experimentation are strategic substitutes, take the cross derivative of the expected utility $\frac{\partial^2 EU_1}{\partial e_{i,2}\partial e_{-i,2}}$:

$$\frac{\partial^2 EU_{i,1}}{\partial e_{i,2}\partial e_{-i,2}} = -\frac{\rho(e_{i,1})\rho(e_{-i,1})Y}{4} < 0$$

For the strategic interaction of an agent's first and her partner's first-stage experimentation, consider the cross derivative of the expected utility $\frac{\partial^2 EU_1}{\partial e_{i,1}\partial e_{-i,1}}$:

$$\begin{aligned}\frac{\partial^2 EU_{i,1}}{\partial e_{i,1}\partial e_{-i,1}} =& -\frac{p^2}{4}Y + \frac{p^2}{4}EU_{i,2} + \frac{\partial^2 EU_{i,2}}{\partial e_{i,1}\partial e_{-i,1}}\left(1 - p \times \frac{e_{i,2}}{2}\right)\left(1 - p \times \frac{e_{-i,2}}{2}\right) \\ & - \frac{\partial EU_{i,2}}{\partial e_{i,1}} \times \frac{p}{2} \times \left(1 - p \times \frac{e_{i,1}}{2}\right) - \frac{\partial EU_{i,2}}{\partial e_{-i,1}} \times \frac{p}{2} \times \left(1 - p \times \frac{e_{-i,1}}{2}\right)\end{aligned}$$

Given the parameters chosen, this gives $\frac{\partial^2 EU_{i,1}}{\partial e_{i,1}\partial e_{-i,1}} < 0$, and an agent's first- and her partner's second-stage experimentation are strategic substitutes. □

Lemma 4. *An agent's second-stage experimentation increases in their partner's first-stage experimentation.*

In the second stage, an agent chooses the optimal experimentation level according to the following first order condition:

$$\frac{\partial EU_{i,2}}{\partial e_{i,2}} = \left(\frac{\rho(e_{i,1})}{2} - \frac{\rho(e_{-i,1})\rho(e_{i,1})}{4}e_{-i,2}\right)Y - c'(e_{i,2}) \stackrel{!}{=} 0$$

For the parameters chosen in the experiment, using that, by symmetry, $e_{i,2} = e_{-i,2}$ in the PBE, this implies that second-stage experimentation in the PBE is given by

$$e^*_{i,2} = \frac{\left(0.9e^2_{-i,1} - 22.71e_{-i,1} + 61\right)e^2_{i,1} + \left(136.28e_{-i,1} - 5.38e^2_{-i,1} - 365.8\right)e_{i,1} + 7.17e^2_{-i,1} - 181.70e_{-i,1} + 487.72}{\left(e^2_{-i,1} - 15.77e_{-i,1} + 39.31\right)e^2_{i,1} + \left(157.24e_{-i,1} - 15.77e^2_{-i,1} - 345.56\right)e_{i,1} + 39.31e^2_{-i,1} - 345.56e_{-i,1} + 722.21}$$

The derivative of $e^*_{i,2}$ defined above w.r.t. $e_{-i,1}$ is positive, so second-stage experimentation is increasing in the partner's first-stage experimentation, if observable. □

Proposition 3. *Experimentation effort is higher in the first stage if this is observable.*

Consider first that effort is observable. An agent will choose first-stage experimentation such that the marginal benefits from experimentation equal the marginal costs from experimentation, given the other player's behaviour according to the following first order condition:

$$\begin{aligned}\frac{\partial EU_{i,1}}{\partial e_{i,1}} =& \frac{2 - pe_{-i,1}}{4} \times pY - c'(e_{i,1}) + \left(1 - p\frac{e_{-i,1}}{2}\right) \times \\ &\left(\left(1 - p\frac{e_{i,1}}{2}\right)\frac{\partial EU_{i,2}}{\partial e_{i,1}} - \frac{p}{2}EU_{i,2}\right) \stackrel{!}{=} 0\end{aligned} \tag{3.7}$$

Consider again the case where effort is not observable. Agents base their decisions on their beliefs about their partner's first-stage effort $\hat{e}_{-i,1}$ instead of the actual partner's effort $e_{-i,1}$ in the second stage. Their partner will base their decisions on their beliefs about their partner's first-stage effort $\hat{e}_{i,1}$. In Eq. 3.7, this affects $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$ and $EU_{i,2}$. I will now consider how these terms depend on the observability of experimentation effort $e_{i,1}$. With observability, $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$ is given by

$$\begin{aligned}\frac{\partial EU_{i,2}}{\partial e_{i,1}} =& \frac{\partial \rho(e_{i,1})}{\partial e_{i,1}}Y\left(\frac{e_{i,2}(e_{i,1}, e_{-i,1})}{2} - \rho(e_{-i,1})\frac{e_{i,2}(e_{i,1}, e_{-i,1})e_{-i,2}(e_{i,1}, e_{-i,1})}{4}\right) \\ &\frac{\rho(e_{i,1})Y}{2}\left(\frac{\partial e_{i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} - \frac{\rho(e_{-i,1})e_{-i,2}}{2}\frac{\partial e_{i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}}\right) + \\ &\frac{\rho(e_{-i,1})Y}{2}\left(\frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} - \frac{\rho(e_{i,1})e_{i,2}}{2}\frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}}\right) - \\ &c'(e_{i,2}(e_{i,1}, e_{-i,1}))\frac{\partial e_{i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}}\end{aligned}$$

Assume now that agents exert experimentation efforts that correspond to the effort levels in the PBE with unobservable effort while $e_{i,1}$ is unobservable. In the unobservable case, $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$ differs from the expression above, as

$$\frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} - \frac{\rho(e_{i,1})e_{i,2}}{2}\frac{\partial e_{-i,2}(e_{i,1}, e_{-i,1})}{\partial e_{i,1}} = 0$$

given that $\frac{\partial e_{-i,2}(e_{i,1},e_{-i,1})}{\partial e_{i,1}} = 0$ with unobservable effort.

Therefore, there is a profitable deviation to exert more effort in the first stage if effort is observable. Recall from Lemma 4 that an agent's second-stage experimentation is increasing in their partner's first-stage experimentation if observable. Furthermore, for a given experimentation level, an agent's second-stage expected utility increases in their partner's second-stage experimentation:

$$\frac{\partial EU_{i,2}}{\partial e_{-i,2}} = \left(\frac{\rho(e_{-i,1})}{2} - \frac{\rho(e_{-i,1})\rho(e_{i,1})}{4} e_{i,2}\right) Y > 0$$

Therefore, effort observability increases incentives to experiment through a change in $\frac{\partial EU_{i,2}}{\partial e_{i,1}}$.

If agents exert experimentation efforts that correspond to the effort levels in the PBE, $EU_{i,2}$ is again constant between the cases with observable and unobservable effort, as posterior beliefs $\rho(e_{i,1})$ and $\rho(e_{-i,1})$ will be the same. Hence, there is a profitable deviation to experiment more with observable experimentation efforts.

Thus, first-stage experimentation efforts in the PBE with observable experimentation are higher than in the PBE with unobservable experimentation. □

Proposition 4. *Experimentation effort is higher in the second stage if this is unobservable.*

In the PBE the agent will maximise her expected utility given her beliefs (sequential rationality). In the case where effort is observable, an agent will choose second-stage effort according to the following condition:

$$\left(\frac{\rho(e_{i,1})}{2} - \rho(e_{i,1})\rho(e_{-i,1})\frac{e_{-i,2}}{4}\right) Y = c'(e_{i,2}) \tag{3.8}$$

If effort is unobservable, second-stage effort will be chosen such that:

$$\left(\frac{\rho(e_{i,1})}{2} - \rho(e_{i,1})\rho(\hat{e}_{-i,1})\frac{e_{-i,2}}{4}\right) Y = c'(e_{i,2}) \tag{3.9}$$

In the PBE, $\hat{e}_{-i,1} = e_{-i,1}$. As $e_{-i,1}$ and $e_{-i,1}$ are higher with observable than with unobservable effort, see Proposition 3:

$$\rho\left(e_{i,1}^{Obs}\right) < \rho\left(e_{i,1}^{Unobs}\right) \quad \wedge \quad \rho\left(e_{-i,1}^{Obs}\right) < \rho\left(\hat{e}_{-i,1}^{Unobs}\right)$$

With $c''(e_{i,2}) > 0$, this implies in the symmetric PBE where $\rho\left(e_{i,1}^{Obs}\right) = \rho\left(e_{-i,1}^{Obs}\right)$ and $\rho\left(e_{i,1}^{Unbs}\right) = \rho\left(\hat{e}_{-i,1}^{Unobs}\right)$ that $e_{i,2}^{Unobs} > e_{i,2}^{Obs}$ such that Eqs. 3.8 and 3.9 hold simultaneously. □

3.C Additional analysis

Tables 3.C.1 and 3.C.2 reproduce the results from Tables 3.1 and 3.2 in the main tax, only including observations from the last 15 rounds of the experimentation game.

		1st stage		
	Unobservable		Observable	
Joint	60.08%	$<^{**}$	71.00%	65.48%
	$\vee^{**}$		$\vee^{***}$	$\vee^{***}$
Separate	48.00%	$<$	55.43%	51.62%
	53.99%	$<^{**}$	63.21%	

Notes: Average experimentation effort in the first stage for experimentation after round 15. Differences in means are tested with Permutation T-tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.1: Experimentation effort per treatment in the first stage in the second half of the experiment

		2nd stage		
	Unobservable		Observable	
Joint	33.04%	$>$	25.34%	29.15%
	$\wedge$		$\wedge^{**}$	$\wedge^{***}$
Separate	37.73%	$<$	38.45%	38.09%
	35.35%	$>$	31.90%	

Notes: Average experimentation effort in the second stage for experimentation after round 15. Differences in means are tested with Permutation T-tests using matching group averages. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.2: Experimentation effort per treatment in the second stage in the second half of the experiment

Table 3.C.3 and Table 3.C.4 reproduce Table 3.4 and Table 3.5 from the main text, respectively, but include subject-level fixed effects. This shows that the fixed effects absorb (parts of) the observed effect of own experimentation on posterior beliefs in treatments with join experimentation, see columns (1) and (2). From this, I can conclude that the variation in experimentation that results in variation of posterior beliefs is mainly between-subject variation in first-stage experimentation. Table 3.C.5 provides estimates of the two-step regression in which residuals are regressed on own first-stage experimentation, controlling for an effect through beliefs. Table 3.C.6 shows the correlation between measures of reciprocity and experimentation behaviour and how this depends on the effort observability. Table 3.C.7 shows how participants' first-stage experimentation correlates with their partners' last-round's experimentation, again depending on the effort observability.

Treatments	*Dependent variable*: Posterior of project's quality			
	Joint projects		Separate projects	
	Observable	Unobservable	Observable	Unobservable
	(1)	(2)	(3)	(4)
Own effort	0.05	-0.05	0.05	-0.02
	(0.06)	(0.05)	(0.04)	(0.04)
Partner's effort	-0.11**	-0.04	0.02	0.03
	(0.05)	(0.05)	(0.03)	(0.03)
Constant	33.19***	38.36***	31.63***	35.65***
	(6.93)	(4.03)	(2.79)	(2.09)
Fixed effects	✓	✓	✓	✓
Observations	2880	2850	2880	2880
Clusters	16	16	16	16
R-squared	0.535	0.593	0.517	0.563

Notes: OLS estimating effect of own and partner's first-stage experimentation effort on posterior of project's quality. (1) and (3) use the partner's actual experimentation effort, (2) and (4) use the subject's belief about the partner's experimentation effort. Robust standard errors (in parentheses) are clustered at the matching group level. Individual fixed-effects included. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.3: The effect of first-stage experimentation on beliefs about partner's posterior

Treatments	*Dependent variable*: Belief about partner's posterior			
	Joint projects		Separate projects	
	Observable	Unobservable	Observable	Unobservable
	(1)	(2)	(3)	(4)
Own effort	-0.00	0.02	0.04*	-0.02
	(0.05)	(0.03)	(0.02)	(0.02)
Partner's effort	-0.09	-0.07	0.08**	0.03
	(0.06)	(0.06)	(0.04)	(0.05)
Constant	35.50***	34.69***	28.36***	34.67***
	(6.57)	(4.35)	(2.48)	(2.34)
Fixed effects	✓	✓	✓	✓
Observations	2880	2850	2880	2880
Clusters	16	16	16	16
R-squared	0.502	0.588	0.503	0.598

Notes: OLS estimating effect of own and partner's first-stage experimentation effort on beliefs about partner's posterior of project's quality. (1) and (3) use the partner's actual experimentation effort, (2) and (4) use the subject's belief about the partner's experimentation effort. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.4: The effect of first-stage experimentation on beliefs about partner's posterior

	Dependent variable: Residuals of 1st-stage regression
Observable	18.15***
	(6.06)
Own effort	0.20**
	(0.07)
Observable × Own effort	-0.28***
	(0.09)
Constant	-12.11**
	(4.62)
Observations	5730
Clusters	32
R-squared	0.030

Notes: OLS estimating difference-in-difference in the effect of own first-stage experimentation between treatments on residuals of regression of second-stage experimentation on posteriors. Elicited beliefs are used for observations from unobservable treatment. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.5: The effect of own first-stage experimentation on second-stage experimentation

	Dependent variable: Experimentation effort			
Treatments	Joint projects		Separate projects	
	(1)	(2)	(3)	(4)
Negative reciprocity	6.49		3.81	
	(4.25)		(5.01)	
Observable	4.34	10.93	2.37	-10.88
	(9.86)	(11.64)	(9.55)	(10.03)
Negative reciprocity × Observable	-9.64		-1.17	
	(6.47)		(6.19)	
Offer		4.31		-2.37
		(4.62)		(3.91)
Offer × Observable		-1.00		11.96**
		(5.94)		(5.70)
Constant	25.87***	55.28***	33.47***	54.06***
	(6.65)	(9.28)	(6.83)	(6.95)
Observations	5730	5730	5760	5760
Clusters	32	32	32	32
R-squared	0.023	0.031	0.004	0.037

Notes: OLS estimating difference in the correlation of behaviour in the ultimatum game and experimentation between treatments. 'Negative reciprocity' refers to the minimum acceptable offer in the ultimatum game, 'Offer' to the amount offered in the ultimatum game. Columns (1) and (3) use second-stage experimentation as the outcome variable, columns (2) and (4) first-stage experimentation. Robust standard errors (in parentheses) are clustered at the matching group level. Robust standard errors (in parentheses) are clustered at the matching group level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.6: Correlation between behavior in the ultimatum game and experimentation

	Dependent variable: Experimentation effort	
	(1)	(2)
Partner's experimentation $t-1$	0.19***	0.04***
	(0.05)	(0.01)
Observable	2.33	
	(5.83)	
Observable × Partner's experimentation $t-1$	0.07	0.04*
	(0.08)	(0.02)
Constant	45.49***	56.29***
	(3.73)	(0.55)
Fixed effects	✗	✓
Observations	11107	11107
Clusters	64	64
R-squared	0.065	0.007

Notes: OLS estimating the differential correlation of partner's first-stage experimentation in the last round on own experimentation depending on whether this was observable. Fixed effects refer to individual-level fixed effects. Robust standard errors (in parentheses) are clustered at the matching group level.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.C.7: Correlation of experimentation with partner's last-round experimentation

3.D Experimental instructions of experimentation game

Instructions part 1

The instructions are simple, and if you follow them carefully, you might earn a considerable amount of money. Your earnings will depend on your decisions and may depend on other participants' decisions.

This experiment consists of two parts. First, we are going to explain part 1 of the experiment to you. After making decisions in part 1, the next part will be explained to you.

In the first part of this experiment, you will repeatedly play a game with changing partners that consists of multiple stages. You will play 30 rounds of this game. Each round you make two decisions. Both your choices and your partner's choices will affect your payoffs.

The task

For each decision you make in each round that you play this game, you receive a budget of €2. Your main task is to decide what share to invest in a project. In each round of the experiment, you will have two opportunities to do so. We will call the percentage share you invest in the project $x\%$.

For investing in this project, you will be charged costs. Costs are higher if you invest a higher share. The higher your investment is, the costlier it becomes to further increase your investment.

More precisely, you can invest between 0% and 100%. If you invest $x\%$, $€2\left(\frac{x}{100}\right)^2$ will be subtracted from your budget.

Examples:

If you invest 0%, the costs are $2 \times \left(\frac{0}{100}\right)^2 = 0€$.
If you invest 30%, the costs are $2 \times \left(\frac{30}{100}\right)^2 = 0.18€$.
If you invest 60%, the costs are $2 \times \left(\frac{60}{100}\right)^2 = 0.72€$.
If you invest 100%, the costs are $2 \times \left(\frac{100}{100}\right)^2 = 2€$.
These costs are subtracted from your budget.

Breakthroughs

The project you can invest in is of high or of low quality. You do not know the quality of the project. With 50% probability, the project is of high quality. With 50% probability, the project is of low quality. This means that if you would face 100 of these projects, you can expect about 50 of these to be high-quality projects.

[*Joint:* You and your partner both invest jointly in the same project. This means that if your partner is investing in a project of high quality, so are you. Similarly, if your partner is investing in a project of low quality, so are you.] [*Separate:* You and your partner invest in separate projects. This means that if your partner is investing in a project of high quality, the project you invest in is not necessarily of high quality, too. Similarly, if your partner is investing in a project of low quality, the project you invest in is not necessarily of low quality, too.]

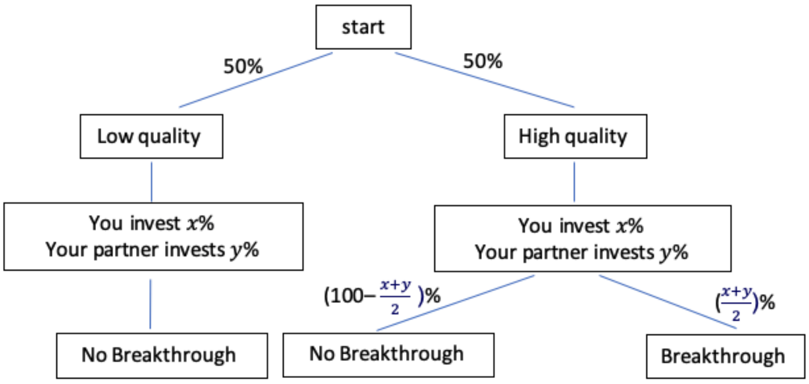
[*Joint:* If your project has a breakthrough, you and your partner each receive a payoff of € 13 from the project's breakthrough.] [*Separate:* Both you and your partner receive a payoff of € 13 each if there is a at least one breakthrough in a project. This breakthrough can be in your project or in your partner's project. If both projects have a breakthrough, you also each receive € 13.]Only high-quality projects can have a breakthrough. Low-quality projects can never have a breakthrough. This means that you will never receive a payoff of € 13 from a low-quality project.

Next to the project's quality, whether there is a breakthrough also depends on how much [*Joint:* you and your partner invest in your joint project.] [*Separate:* is invested in each of the projects. The more either of you invests in his or her project, the more likely that project has a breakthrough.] If you face a high-quality project, the probability of a breakthrough increases with the share [*Joint:* you and your partner together invest.] [*Separate:* that is invested in the project.] More specifically, the probability of a breakthrough is [*Joint:*the average of your and your partner's investment share.] [*Separate:* half of the investment share.] If you invest x% [*Joint:*and your partner invests y%,] the probability of a breakthrough [*Joint:* is thus $\frac{x+y}{2}$%] [*Separate:* in your project is $\frac{x}{2}$%] for high-quality projects. [*Separate:* If your partner invests y%, the probability of a breakthrough in your partner's project if $\frac{u}{2}$% it is a high-quality project.]

If [*Joint:*you are] [*Separate:* someone is] facing a high-quality project and [*Joint:* both you and your partner invest] [*Separate:* that person invests] a share of 100% in [*Joint:* this] [*Separate:* his or her] project, [*Joint:*you will certainly have a breakthrough and will both receive € 13.] [*Separate:* there will be a breakthrough with a probability of 50%.] On the other hand, if [*Joint:*both you and your partner invest nothing in this project, you will never have a breakthrough,] [*Separate:* someone invests nothing in his or her project, there will never be a breakthrough], no matter whether the project is of high quality or not.

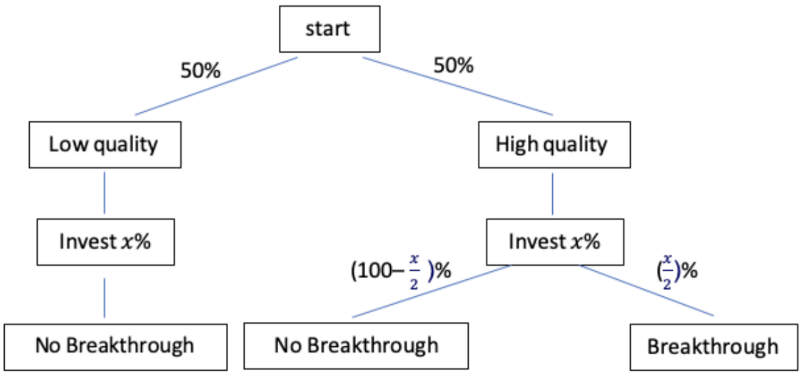
You can determine the likelihood of a breakthrough in a project as follows.

[*Joint:*



]

[*Separate:*



]

After you and your partner have made your investment choices, the computer will determine whether there actually is a breakthrough [*Separate:* for both projects]. To determine this, the computer will use the breakthrough probability.

Examples:

Let's say that you invest 20% [*Separate:*in your project] and your partner invests 64% in [*Joint:*the] [*Separate:*his or her] project. [*Joint:*If the project is of low quality, there will not be a breakthrough. If the project is of high quality, the probability of a breakthrough is $\frac{20\%+64\%}{2} = 42\%$.] [*Separate:* If your project is of low quality, there will not be a breakthrough for your project. If your partner's project is of low quality, there will not be a breakthrough for his or her project. If your project is of high quality, the probability of a breakthrough in this project is $\frac{20}{2} = 10\%$. If your partner's project is of high quality, the probability of a breakthrough in this project is $\frac{64}{2} = 32\%$]

There can only be a breakthrough after the first or after the second investment decision. If [*Joint:*the] [*Separate:*any] project has a breakthrough after your first investment decision, you cannot invest anymore. You will receive € 13 plus the budget of your second investment decision.

Within one round, the project you are investing in does not change. If the project is of high quality for your first investment decision, it will also be of high quality for your second investment decision in this round. Similarly, if the project is of low quality for your first investment decision, it will also be of low quality for your second investment decision. This means that the first investments and results of the first investments may contain information relevant to your second decisions.

In each new round you will face a new project. While in each round you face a project of high quality with 50%, the project's actual quality in one round does not say anything about the project's quality in any other round.

Your decisions

After being matched with a partner for a round, you will be asked to make three types of decisions. First, an investment decision, second, predictions about the project's quality and your partner's choices, and third, another investment decision. We now describe each of these three decisions in more detail.

First investment decision

For your first investment decision, you decide which share you want to invest and then submit your decision. You can use an on-screen calculator that will allow you to calculate your costs for any given investment and the probability of a breakthrough. You will see this interface on a later screen.

If there is a breakthrough after the first investment decisions, you will receive the payoff minus your costs from the first investment decision added to your budget. The breakthrough terminates this round. The second investment decisions are in this case not relevant.

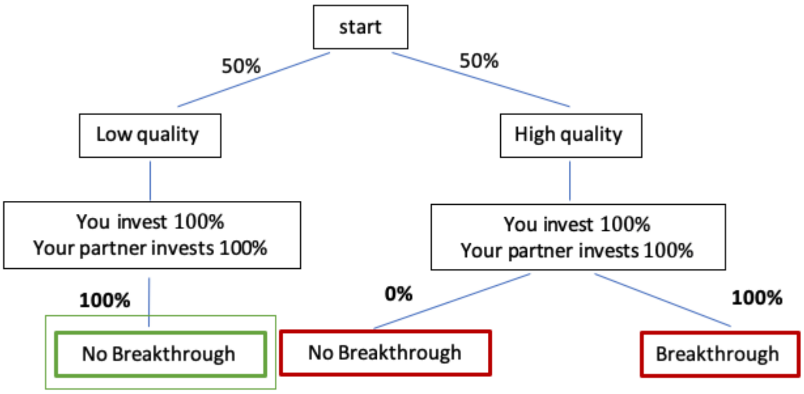
Prediction task

After the first investment decision, you are asked to state your beliefs about the project's quality, your partner's investment as well as your partner's beliefs about the project's quality. You will also be paid according to your performance in this task. This task will be explained in more detail on a later screen.

Second investment decision

After your first investment decision, we will ask you how much you would want to invest in the project **if there was no breakthrough after the first investment deci-**

sions. Your second investment decision will then only be implemented in case there was indeed no breakthrough after the first investment decisions. This means that you should decide how much you think is best to invest in a project where there has not yet been a breakthrough. For the case that there was a breakthrough after the first investment decisions, you receive your €2 second-period budget added to your payoff. So, after a breakthrough you will still be asked to make the second investment decision, but this will only be relevant for your payoffs if there indeed was no breakthrough! Decide as if there was no breakthrough so far. [*Joint:*Realize that if you would know for sure that both you and your partner had invested your entire budgets in the project in the first investment decision, while there was no breakthrough, then, the project cannot be of high quality. This is why: the probability of a breakthrough if the project is of high quality and you both invest 100% is given by $\frac{100\%+100\%}{2} = 100\%$. This means that if you observe no breakthrough, you are for sure in the far-left green branch of the tree below. You would have certainly seen a breakthrough if the project were of high quality.

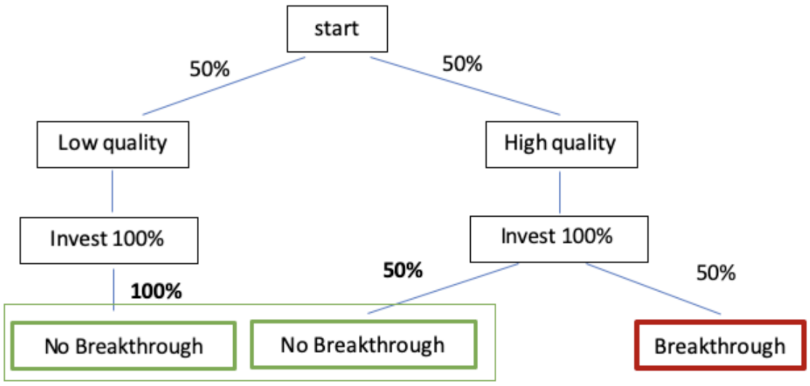


]

[*Separate:* Realize that if someone had invested his or her entire budget in his or her project in the first investment decision, while there was no breakthrough, then it is twice as likely to have a low-quality project than a high-quality project. This is why: You see below that we must be in one of the two left branches of the tree, within the green box. The project could be of low quality, then we would observe no breakthrough with a probability of 100% (far left branch). Alternatively, the project is of high quality, but there was no breakthrough (middle branch). This is only half as likely, as if you invest 100% and the project is of high quality, the probability of no breakthrough is only $\frac{100\%}{2} = 50\%$.

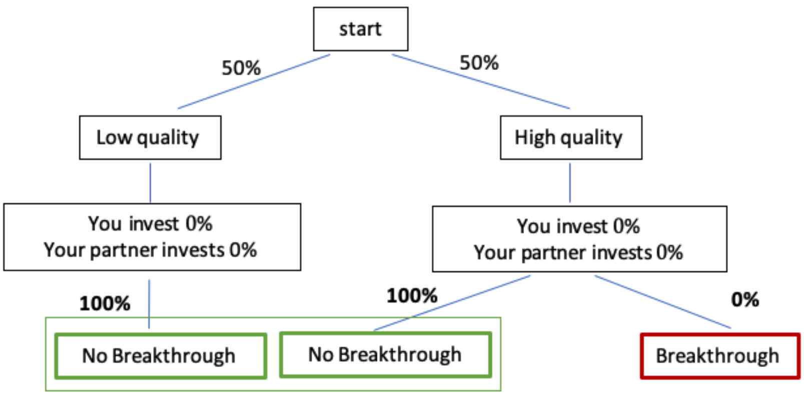
]

If, in contrast, [*Joint:* both you and your partner invest] [*Separate:* someone invests]nothing in the project in the first investment decision, then you cannot learn anything new about the project. The probability of a breakthrough is 0%. [*Joint:* So if there is no breakthrough, it is equally likely to be in one of the two green branches

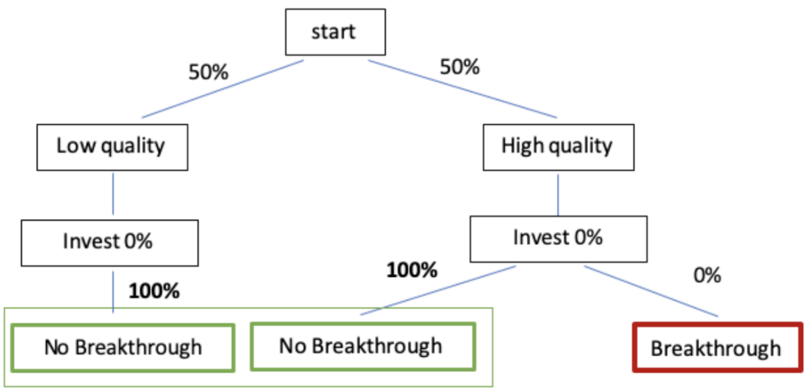


below, the far left or the middle one.] The probability that the project is of high quality is in this case still 50%.[*Separate:* Below you see that no matter whether the project is of high or low quality, the probability of no breakthrough is always 100%. It is then equally likely to be in the left low-quality branch or in the middle high-quality branch.]

[*Joint:*



] [*Separate:*



]

We will show you a graph to illustrate your expected payoff from making specific investments. As before, you can also use on-screen calculators that will allow you to calculate your costs for any given investment and the probability of a breakthrough.

Now, this depends on your beliefs about the probability that the project is of high quality. This will be illustrated on a later screen.

Feedback

After your first investment decision, you will [*Unobservable:* not] see which share your partner invested. Your partner will also [*Unobservable:* not] see which share you invested. After the second investment decision, you will see whether there was a breakthrough and how much your payoff from this round is.

Your partner

Your partner is anonymous and so are you. Your partner is the same for both investment decisions. You face the same decision situation. After each round of the experiment, you will be randomly assigned to a new partner. We ensure that you are never linked to the same partner for two rounds in a row. Also, your actions in any round have no influence on anything that happens in other rounds and are not known to your partners in following rounds.

Payoffs from this task

To summarize, your payoffs from each investment decision are the following:

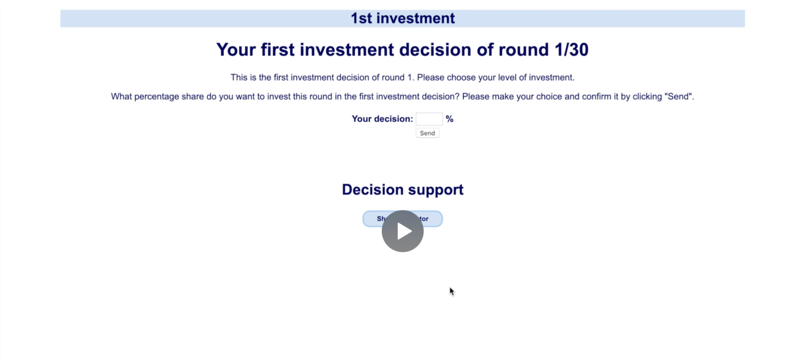
- If you invest x_{1st}% in this project in the 1st investment decision:
  - If you [*Separate:*, your partner or both of you] achieve a breakthrough: $2+13-2\times\left(\frac{x_{1st}}{100}\right)^2$
  - If [*Joint:* you do not achieve] [*Separate:*, no one achieves] a breakthrough: $2-2\times\left(\frac{x_{1st}}{100}\right)^2$
- If you invest x_{2nd}% in this project in the 1st investment decision:
  - If there was no breakthrough after the first investment decision:
    - If you [*Separate:*, your partner or both of you] achieve a breakthrough: $2+13-2\times\left(\frac{x_{2nd}}{100}\right)^2$
    - If [*Joint:* you do not achieve] [*Separate:*, no one achieves] a breakthrough: $2-2\times\left(\frac{x_{2nd}}{100}\right)^2$
  - If there was a breakthrough after the first investment decision:
    - 2

At the end of the experiment, **two rounds** of this investment game will be randomly chosen by the computer for payment. Every round is equally likely to be chosen for payment, so your actions in a round have no influence on whether that round will be paid out. Aside from the payment from the investment game, two different rounds will be chosen from which the prediction task will be paid. We will explain on another screen how payment is determined for the prediction task.

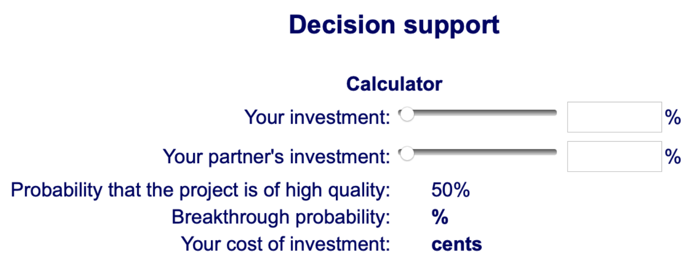
Illustration

First investment decision

For your first investment decision, you make a choice about which share to invest in the project. Please see this video illustrating how to make your first investment decision. You can pause the video at any moment, re-watch it as many times as you like and put it on full-screen if you prefer this. Please try using the calculator below.

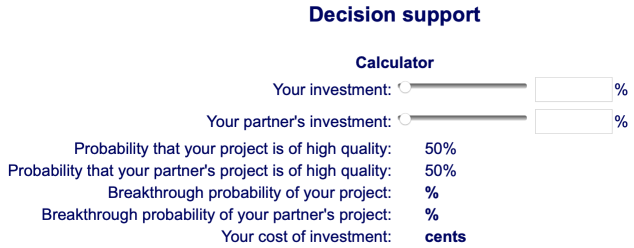


[*Joint:*



]

[*Separate:*



]

Second investment decision

For the second investment decision, you again make a choice about which share of your budget to invest in case the project did not have a breakthrough after the first investment decisions.

On your decision screen, you will have the opportunity to see a graph of your expected payoff from investing a certain share. Expected payoff means that this is not a certain payoff from investing this share, but that this is what you are going to receive in expectation. If you would do this investment frequently, on average you would get the expected payoff. The realized payoff from investing a share x will always be either $13 - 2 * (\frac{x}{100})^2$ (if there is a breakthrough) or $-2 * (\frac{x}{100})^2$ (if there is no breakthrough), which will be added to your budget of € 2.

Your expected payoff depends on [*Joint:* three] [*Separate:*four] factors:

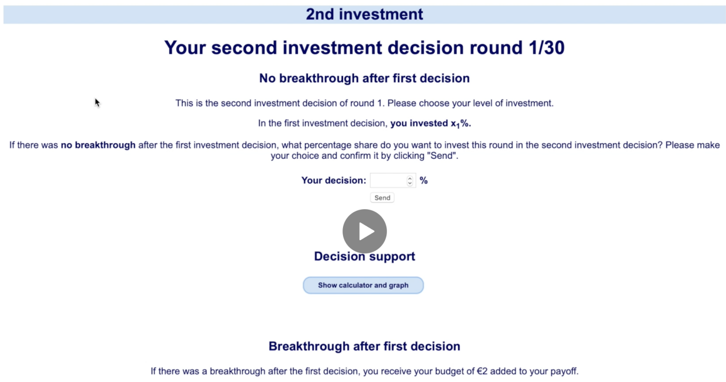
1. How likely [*Joint:* the] [*Separate:* your] project is of high quality: Your expected payoff is higher if [*Joint:* the] [*Separate:* your] project is more likely of high quality, as only [*Joint:* high-quality projects can result in a breakthrough and thus in a payoff of € 13 for you and your partner.] [*Separate:* then your project can have a breakthrough.]

2. [*Separate:* How likely your partner's project is of high quality: Your expected payoff is higher if your partner's project is more likely of high quality, as only then your partner's project can have a breakthrough.]

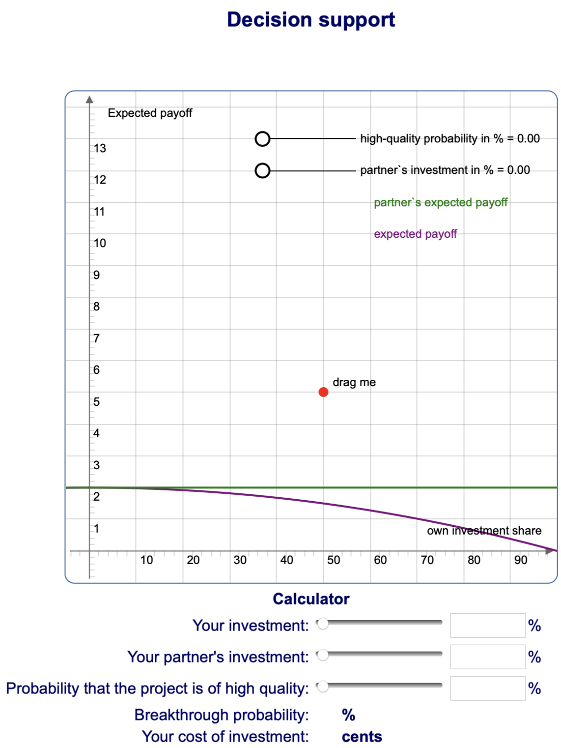
2/3 The share you invest: This also increases the probability of a breakthrough [*Separate:* in your project] if the project is of high quality.

3/4 The share your partner invests: If he or she invests more, this increases the probability of a breakthrough[*Joint:* if the project is of high quality.] [*Separate:* in his or her project if this project is of high quality.]

Please see this video illustrating how to make your second investment decision. You can pause the video at any moment, re-watch it as many times as you like and put it on full-screen if you prefer this.

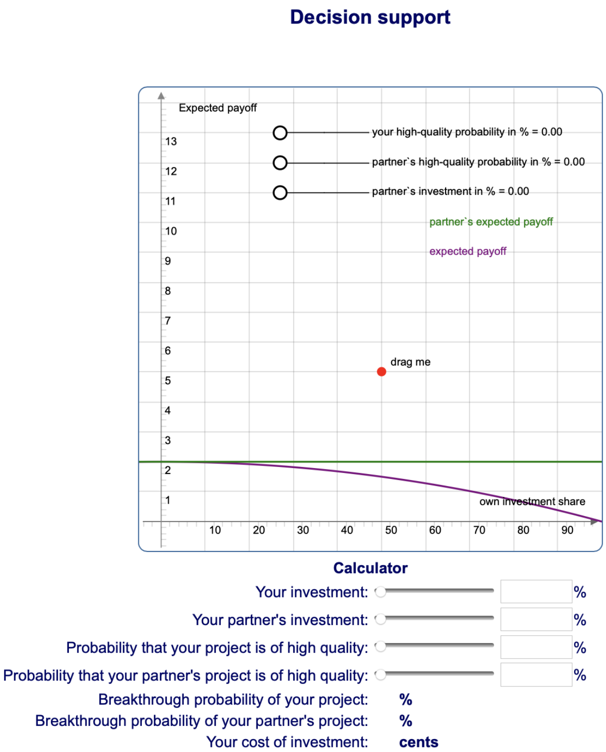


Please try using the calculator and the graph below.
[*Separate:*



]

[*Joint:*



]

Summary

- You and your partner can [*Joint:* invest in a common project] [*Separate:* each invest in a project]

- [*Joint:* This] [*Separate:* A] project is of high quality with probability 50% and of low quality with probability 50%

- [*Joint:* You don't know whether the project is of high or of low quality] [*Separate:* You know neither whether the project you face nor whether the project your partner faces is of high or of low quality]

- You and your partner invest in [*Joint:* the same project with the same quality] [*Separate:* separate projects which can have different qualities]

- You have a budget of €2 for each investment decision

- Investing a share of x% in the project costs $2 \times \left(\frac{x}{100}\right)^2$

- You and your partner receive a payoff of €13 if there is [*Joint:* a breakthrough] [*Separate:* at least one breakthrough in one of the two projects.]

- If the project is of high quality, the probability of a breakthrough [*Joint:* is given by the average of your investment share x% and your partner's investment share y%, $\frac{x+y}{2}$%.] [*Separate:* in this project is given by half the share of x% which you or your partner invested in that project: $\frac{x}{2}$%]
- If [*Joint:* the] [*Separate:* a] project is of low quality, no breakthrough is possible
- Breakthroughs end the project
- If there was no breakthrough after the first investment decision, you and your partner can invest again in [*Joint:* the project] [*Separate:* your projects]
- After the first investment decision, you will [*Unobservable:* not] see which share your partner invested. Your partner does [*Unobservable:* not] see the share you invested.
- Your second investment decision is for the case that there was no breakthrough after the first investment decision
- Your partner and the project are the same for the first and second investment decision, but change every round

3.E Experimental instructions of belief elicitation[15]

Your predictions

In each round, you will make several guesses after the first decision of the investment game.

1. [*Unobservable:* What share (in percent) did your partner invest in the first investment decision of this round?]
2. What is the probability (in percent) that [*Joint:* the] [*Separate:* your] project is of high quality if there was no breakthrough after the first investment decision?
3. What does your partner think is the probability (in percent) that [*Joint:* the] [*Separate:* his or her] project is of high quality if there was no breakthrough after the first investment decision?
4. What share (in percent) will your partner invest in the second investment decision of this round if there was no breakthrough after the first investment decision?

[15]These instructions are built on and in parts taken from instructions in Babcock et al. (2017)

Before you give your guesses, we will remind you of the share you invested in the first investment decision.

Your guess will secure you a payment of either € 2 or € 0 for each guess. If you win, you receive € 2. If you lose, you instead receive € 0 for your guess. Your payoffs from this task are such that you maximize the probability of receiving a prize of € 2 by stating your best guess for each question.

Except for guess #1, you can see that you are asked about your beliefs in case there was no breakthrough after the first investment decision. Therefore, the computer randomly picks two rounds in which there was no breakthrough after the first investment decision for payment. From these two randomly selected rounds, you have the chance to win € 2 for each of the guesses depending on your answer. So you can earn up to € [*Unobservable:* 16] [*Observable:* 12] from your guesses.

To determine your probability of winning the prize for each guess, we will compare your guess to what actually happens. We designed the payment rule such that you can secure the largest chance of winning the prize by reporting your most-accurate guess. Below, you can read more about how we determine whether you win the prize. To maximize your chances of winning the prize, it is not necessary that you understand how this works. While the mechanism may look complicated, what it means for you is simple: you have the highest chance of winning € 2 if you report your best guess for each question.

Click here for more information on the mechanism

For each question, we will use your guess to calculate a chance-to-win. How we do this is explained below. We use this chance-to-win to determine whether you win € 2. The computer generates a random number between 1 and 100 separately for each question. Each of the numbers is equally likely. You win € 2 if this random number equals or falls below your chance-to-win, and you earn € 0 if the random number exceeds your chance-to-win.

To maximize your earnings, you should submit a guess that secures a high chance-to-win for the events you think are most likely, and a low chance-to-win for the events that you think are least likely. If you, for instance, believe that it is very likely that the project is of high quality, you should submit a guess that secures a high chance-to-win for the case that the project is of high quality.

To secure that it is in your best interest to enter your best guess, we use the following procedure to calculate your chance-you-win for guess #2: Suppose you submitted a guess of p_1 that [*Joint:* the] [*Separate:* your] project is of high quality. Then your chance-to-win depends on whether the realized quality of [*Joint:* the] [*Separate:* your] project is high or low. If the project is of high quality, your chance-to-win is given by the equation:

$$Chance-to-win: \left(1-\left(1-\frac{p_1}{100}\right)^2\right)\times 100$$

If the project is of low quality, your chance-to-win would be given by the equation:

$$Chance-to-win: \left(1-\left(\frac{p_1}{100}\right)^2\right)\times 100$$

This means that you have the highest probability of earning the €2 if your guess p_1 is what you believe is the probability that [*Joint:* the] [*Separate:* your] project is of high quality.

Example

Let's say that your best guess of the probability that [*Joint:* the] [*Separate:* your] project is of high quality is 40%. If you state this truthfully, then your chance-to-win is $\left(1-\left(1-\frac{40}{100}\right)^2\right)\times 100 = 64$ if [*Joint:* the] [*Separate:* your] project is of high quality and $\left(1-\left(\frac{40}{100}\right)^2\right)\times 100 = 84$ if [*Joint:* the] [*Separate:* your] project is of low quality. As your best guess of the probability that the project is of high quality is 40%, the probability that you receive the prize is then 64*0.4+84*0.6=76%. If, for instance you would untruthfully state that your best guess of the probability is 70%, the probability of receiving the prize is lower: Your chance-to-win is $\left(1-\left(1-\frac{70}{100}\right)^2\right)\times 100 = 91$ if [*Joint:* the] [*Separate:* your] project is of high quality and $\left(1-\left(\frac{70}{100}\right)^2\right)\times 100 = 64$ if [*Joint:* the] [*Separate:* your] project is of low quality. As your best guess is 40%, the probability that you receive the prize is then 91*0.4+51*0.6=67%, which is lower.

For the remaining questions, we use the following procedure: Suppose that you submitted a guess that your partner's belief about the probability that [*Joint:* the] [*Separate:* his or her] project is of high quality is x% (guess #3) or that your partner is going to invest [*Unobservable:* (or invested)] x% in the second investment decision of the round (guess #4 [*Unobservable:* and guess #1]). Then your chance-to-win depends on what your partner actually believes is the probability of [*Joint:* the] [*Separate:* his or her] project being of high quality or on how much he or she actually invested in that second investment decision. Let's call either of these percentages y%. Your chance-to-win will then be given by the equation:

$$Chance-to-win: \left(1-\left(\frac{y}{100}-\frac{x}{100}\right)^2\right)\times 100$$

This means that you have the highest probability of earning the €2 if your guess x is what you believe is your partner's belief (guess #3) or what you believe he or she will invest (or invested) (guess #4 [*Unobservable:* and guess #1]).

CHAPTER

4

PITFALLS OF PAY TRANSPARENCY: EVIDENCE FROM THE LAB AND THE FIELD.

This chapter is based on Brütt and Yuan (2022).

4.1 Introduction

Despite advances in promoting equal pay, wage discrepancies between men and women still characterize the overwhelming majority of labor markets (Blau and Kahn, 2017). The EU-wide gender pay gap amounted to 13% in 2020.[1] In an effort to close this persistent gap, the European Commission has recently become more vocal in urging all member states to adopt pay transparency legislation and makes EU-wide wage transparency regulation a political priority (European Commission, 2021). Several EU countries, as well as multiple states in the U.S., have already adopted a variety of measures against pay secrecy.[2] These measures include instruments such as pay information for job seekers, the right to access pay information for workers in similar positions, and company-level gender pay gap reporting duties. In this paper, we examine the impact of a particular wage transparency measure introduced in Germany in 2017 and adopt an online experiment to explore potential mechanisms that determine the impact of such regulation.

Wage transparency regulations target wage negotiations by aiding employees in bargaining for a wage they deem fair (see Recalde and Vesterlund, 2022, for a recent overview). Gender differences in negotiation outcomes are deemed one possible source for the remaining pay gap. Wage transparency policies are instruments designed on the one hand to reveal discriminatory practices and on the other hand to correct misguided beliefs about co-workers' wages. We will focus on the latter component of wage transparency. Wage information reduces the informational asymmetry between workers and firms, which may prove advantageous in negotiations for both men and women. Since women tend to have more pessimistic beliefs about average and future wages and there is a substantial gender difference in earnings expectations (see e.g. Kiessling et al., 2019; Briel et al., 2022; Boneva et al., 2022), wage information may prove particularly beneficial to women and thus contribute to a reduction in the gender pay gap. This study will investigate the interaction of the correction of misspecified beliefs and wage negotiations.

We combine field and laboratory data to address the empirical success of current wage transparency laws in Germany and study the requirements for an effective wage transparency regulation. As a first step, we study the effect of the transparency law on wages in Germany. The laboratory experiment expands on this. It considers how a key feature of the German legislation, the fact that information is available only on request, may limit the usefulness of such regulation. Furthermore, we examine whether wage transparency works differently in environments that also allow for performance

[1] Source: Eurostat, 2021

[2] For the EU, see the fact sheet of the European Commission on pay transparency measures across the EU (European Commission, accessed December 2021). See the Women's Bureau of the U.S. Department of Labor for an overview of state measures for pay transparency (U.S. Department of Labor, accessed December 2021)

comparison. We study this in the laboratory, as there is no naturally occurring exogenous variation in the types of transparency regulation.

While the German transparency law allows both men and women to request wage information, the law has the explicit goal of reducing the wage gap between men and women for comparable activities. It gives employees in firms with more than 200 employees the right to request information about the compensation that comparable workers receive. Leveraging German administrative employer-employee matched data, our identification strategy exploits variation in the transparency policy based on firm size and over time. We employ both a difference-in-difference analysis and a difference-in-discontinuities analysis to provide a quasi-experimental evaluation of the impact of wage transparency on the gender pay gap.

EU countries have introduced a large set of heterogeneous wage transparency measures. The German wage transparency regulation lends itself particularly well to studying the effect of wage transparency on negotiations. In contrast to wage transparency regulation introduced in many other countries,[3] the German regulation permits workers to ask for wage information of the median worker with comparable work. Compared to wage statistics aggregated at a higher level, this information appears more relevant for wage negotiations. As the work, by definition, is comparable, it allows workers to argue for comparable compensation.

We do not find any evidence that the introduced wage transparency regulation decreased the gender pay gap in Germany. Both wages of men and women are unaffected and this finding is robust and independent of the specification we consider. We can estimate this null effect with high precision. In our preferred specification of the difference-in-difference analysis, we can exclude in our 95% confidence interval that the treatment effect of the introduction of the wage transparency law is larger than a 1.29 percentage point reduction in the gender pay gap, with a point estimate smaller than 0.1 percentage points. This result remains qualitatively the same for subgroups which may be affected to the largest extent due to their unionization status. Moreover, the regulation also does not push employees to move to a different employer.

Given this result, we set out to better understand the determinants of when and how wage transparency measures can deliver on their promise to reduce gender differences in wage negotiations. We do so theoretically and experimentally. In a simple theoretical model, we propose a novel mechanism that captures the impact of wage and performance information as an information shock that corrects misspecified beliefs. Since both men and women can request wage information, it is ex-ante not clear that women benefit more from transparency measures. Our model shows how the provision of both wage and performance information can decrease the gender pay gap in a Nash bargaining framework. We assume the worker cares about receiving

[3]See Section 4.2 for a discussion of studies analyzing different types of wage transparency measures.

a wage he or she perceives as fair. We formalize this as a preference for receiving a piece rate similar to the worker's beliefs about the piece rates of comparable workers.

In our experiment, we address potential barriers to the effectiveness of the type of wage transparency policies currently implemented in Germany. In the experiment, workers and firms negotiate bilaterally over the split of resources they have produced in a task. Between treatments, we vary the information provided to the worker. As a first barrier, we consider whether requiring that wage information is actively requested diminishes the potency of this type of intervention. In Germany, only 4% of eligible employees had requested wage comparison a year after the implementation of the wage transparency regulation.[4] To analyze whether automatic access to wage information can increase its effectiveness, our experiment varies whether wage information is absent, provided upon request for a small fee, or exogenously provided. If wage information is available, the experiment informs workers of the wage of a worker who was previously paired with the same firm and did the same task.

Second, we examine the type of environments that facilitate the use of wage information. We argue that wage information is particularly useful in settings where employees are aware of their relative performance. Think of settings where performance is easily measurable and information on this is accessible, such as in sales departments, compared to a setting where it is less well observable, such as in HR departments. In the absence of performance information, employees cannot evaluate whether wage differences are due to differences in performance. This, however, may be crucial information when bargaining. Therefore, we hypothesize that the joint provision of wage and performance information has the strongest effect on gender wage differences.

In line with our findings from the field, our experimental results show that workers do not earn significantly more if wage information is provided endogenously. However, we show that workers obtain a higher wage if wage information is provided exogenously. Removing the barrier to wage information thus helps workers overall. This effect is not gender specific. Changing beliefs about wages, therefore, does not narrow the gender pay gap. Similarly, the provision of performance information increases workers' wages. Workers' wages mirror performance differences more closely if these are observable, resulting in a reduced variance in piece rates between workers. The effect of performance information is, however, also not different between men and women. These findings suggest that decreasing the informational asymmetry between worker and firm increases the workers' bargaining power. In our setting, this increase in bargaining power is not larger for women than for men.

Our experiment also shows that different types of wage transparency regulation

[4]See Report by the Federal Government on the effectiveness of the Act to Promote Transparency in Wage Structures among Women and Men (Germany Federal Ministry for Family Affairs, Senior Citizens, Elderly and Youth, accessed July 2022)

can have unintended consequences. First, we observe that if wage information is provided on request only, employees requesting this information receive lower wages than employees being provided with this information exogenously. Second, receiving wage information reduces women's propensity to enter negotiations. While the share of decisions to opt out of negotiations in our experiment is low enough such that this does not translate into a significant change in the wage gap, opting out of wage negotiations is associated with a substantial expected wage loss. Hence, wage transparency regulation might also backfire by deterring women from negotiating at all.

The effects we find in the laboratory are small and have to be treated with caution. Wage transparency policy in the field does not only serve the purpose of correcting beliefs about other workers' wages. Instead, an employer's discriminatory behavior and wage disparities are made apparent. In our laboratory experiment, we abstract from these aspects. Nevertheless, our results have policy implications for the design of wage transparency regulations. We underline the importance of studying the distinct features of wage transparency regulations before rolling out future policies. First, the analyzed 'pay information right' regulation, which allows employees to request wage information, has so far not been successful. 'Pay reporting duties', which require employers to provide this information, might fare better (Bennedsen et al., 2022; Duchini et al., 2020). Advantages of providing pay information to everyone have, however, to be weighed against potential downsides, such as possibly lower job satisfaction (Card et al., 2012). Second, policymakers may need to consider distinct wage transparency policies depending on their specific goal. We see in our experiment that workers overall might benefit from transparency, but this does not reduce the gender wage gap. As women are deterred from entering negotiations by wage transparency, potentially due to the social comparison it entails, providing wage information may have adverse effects.

This paper proceeds as follows: Section 4.2 reviews the related literature. Section 4.3 provides an overview of the institutional setting and the analysis of the field data examining the effects of the German wage transparency law. We turn to the experiment in Section 4.4, first explaining our theoretical predictions, then the experimental design and the results from the experiment. Section 4.5 briefly concludes.

4.2 Related literature

Our results aim to contribute to two strands of the literature. First, we contribute to the growing literature on the impact of wage transparency laws in different settings. So far, no consensus has been reached on the effects of transparency measures.

The literature shows that wage information significantly reduced the gender pay gap among academics in Canadian and British universities (Baker et al., forthcoming; Gamage et al., 2020). A particular focus has so far been on the study of 'pay reporting duties', where companies are required to disclose gender-specific wage statistics. These policies are often implemented based on a size threshold and only affect firms with sufficiently many employees, an assignment rule that has been exploited in other studies. Such a reform in the U.K. resulted in more women being hired in above-median-wage jobs and a reduction in the male hourly wages (Duchini et al., 2020). The reform resulted in a decrease in the gender pay gap (Blundell, 2021). These findings are in line with evidence from Denmark, where slower wage growth for men drove a significant decrease in pay inequality (Bennedsen et al., 2022).

There are, however, not only success stories of wage transparency regulations. Publicly disclosed wages reduced the managers' compensation in California (Mas, 2017) and wage transparency can reduce job satisfaction (Card et al., 2012). More closely related to our research, the Austrian Pay Transparency Law did not impact wages (Gulyas et al., forthcoming; Böheim and Gust, 2021). Wage information in Austrian job advertisements also did not affect gender sorting into better-paid jobs (Bamieh and Ziegler, 2022). Greater transparency in the U.S. private sector has even reduced overall wages (Cullen and Pakzad-Hurson, 2021).

Our study contributes another data point to the conflicting results in this growing literature. Our aim, however, is broader than this. So far, there is no evidence on what could make a transparency law effective. One contribution of this study is to investigate the unique transparency policy implemented in Germany that mandates the provision of wage information of co-workers in comparable positions on request, rather than the publication of firm-wide wage averages. Therefore, we do not study transparency measures classified as 'pay reporting duties', but a different class of measures coined 'pay information rights'. We analyze this endogeneity of receiving wage information more closely in our experiment. Furthermore, the information on wages paid to workers in similar positions could plausibly be more useful in wage negotiations than aggregate wage statistics. Therefore, our contribution is to investigate a setting in which wage information particularly lends itself to be used in negotiations.

The second strand of literature we contribute to is the experimental literature that studies gender differences in negotiations. Wage negotiations are seen as one source of the gender pay gap. Women enter negotiations less often, ask for lower wages (Roussille, 2020), and, depending on the exact setting, receive worse negotiation outcomes, see e.g. Bowles et al. (2005), Azmat and Petrongolo (2014), Mazei et al. (2015), Hernandez-Arenaz and Iriberri (2018) or Recalde and Vesterlund (2022) for overviews. In particular, settings with high ambiguity over the possibility to negotiate, that are competitive, and in which women have to negotiate on behalf of themselves (Bowles et al., 2005; Amanatullah and Tinsley, 2013) are prone to result in lower wages for women. Field evidence is in line with these findings. Flexible wage policies that allow for wage bargaining increase the gender wage gap among public school teachers (Biasi and Sarsons, 2022); women have a lower propensity to enter negotiations (Greig, 2008), especially if there is ambiguity (Leibbrandt and List, 2015); and female graduates request lower wages in their starting-wage negotiations (Säve-Söderbergh, 2019).

Closest to our work is the literature that considers how information and interventions in negotiations affect gender differences. One possible intervention is to force women to negotiate more. Laboratory evidence, however, suggests that this does not benefit women. If women are forced to enter negotiations, they have to enter negotiations that are not profitable (Exley et al., 2020). The other extreme would be a negotiation ban, which appears to be more successful. Banning negotiations reduces inequalities between men and women in an experiment (Gihleb et al., 2020).

There is a small literature explicitly focusing on transparency interventions in negotiations. The literature shows that providing wage information can affect employees' behavior. In a field experiment, employees exert more effort if they find out that their managers earn more than expected (Cullen and Perez-Truglia, 2022). There is no evidence of any gender-specific impacts of this information. Some laboratory studies consider the effect of the provision of social information on wage negotiations. Focusing on the dynamic response of firms to the requirement of providing wage information, recent evidence points to higher and more equal wage offers with exogenous compared to endogenous information (Werner, 2019). In contrast to our study, Werner (2019) does not study gender-specific effects and focuses on firm behavior. In an ultimatum bargaining experiment that varies whether information on previous pay requests and average offers are provided, the gender gap in negotiated wages disappears if information is available (Rigdon, 2012). In contrast to our study, the information provided here stems from male participants only.

We add to this strand of literature by examining both the difference between endogenous and exogenous information provision and the interaction of wage and performance information. Furthermore, we focus on the effect of information on gender wage differences and take a closer look at the mechanisms that drive the effect of information provision by studying how beliefs are corrected. Specifically, we capture the role of confidence and beliefs about others' wages.

4.3 Field data

In this section, we will first introduce the institutional setting relevant for the wage transparency law in Section 4.3.1, then describe the data used in our analysis in Section 4.3.2, explain our identification strategy in Section 4.3.3 and finally discuss our results in Section 4.3.4. We provide robustness checks in Section 4.3.5.

4.3.1 Institutional setting

Germany has one of the largest gender wage gaps in the EU, with women earning on average 18.3% less than men in 2020.[5] In March 2017, the German federal parliament passed new legislation to battle gender-based wage inequality. This legislation was adopted in June of that year as the 'Gesetz zur Förderung der Transparenz von Entgeltstrukturen' (BGBl. I S. 2152, referred to here as 'wage transparency law'). The goal of this law is to eliminate inequalities across gender in wages for the same work. This law includes several instruments that are in place to enforce this ban of unequal pay. We focus on the pay information rights that are part of this law, which came into effect on January 06, 2018.

The pay information rights prescribe that employees in establishments with more than 200 employees working for the same employer can request information about the median wage of an employee of the opposite gender doing comparable work. This comparison group has to comprise at least six individuals to prompt the provision of wage comparison. The request will be handled by the works council or the employer itself.[6] Employees can use this right every two years or more frequently if working conditions substantially change.

The German wage transparency regulation differs in several aspects from wage transparency regulations implemented in other countries. First, workers have to actively ask their employer or works council to provide the information ('pay information right'). This is in contrast to transparency regulation implemented in e.g. Denmark, the U.K. or Austria ('pay reporting duties'). Second, employees receive a different type of information than in several other countries. Instead of receiving wage statistics that are aggregated at the company level, such as in Austria or the U.K., the employee can request wage information on a worker in a comparable position. This second point makes this transparency regulation particularly interesting to study in relation to wage negotiations; in contrast to company-wide wage statistics, wage information of an employee with a comparable task is an instrument that allows women to argue for a comparable wage.

On the one hand, there is some anecdotal evidence that this regulation has an impact on women's wages. For instance, a female head of department won a discrimination lawsuit in the Federal labor Court using information obtained through the wage transparency law.[7] On the other hand, survey data point to low uptake among

[5] Source: Eurostat, 2022

[6] This depends on whether employers are bound to collective bargaining agreements and on whether a works council exists.

[7] Source: Deutsche Welle, 2021

employees in eligible firms (cf. fn 4). So far, no thorough analysis of the overall effects of this regulation exists.

4.3.2 Data description

Our primary data source stems from the German Institute for Employment Research (IAB). We utilize the 'Linked-Employer-Employee-Data from the IAB' (LIAB). This employer-employee matched data set combines administrative data with an annual establishment survey. We observe the complete employment histories of 1,688,101 employees at firms surveyed in the IAB Establishment Panel, a representative sample of nearly 15,500 German establishments.

Our primary analysis will use only the administrative data on individuals and establishments from LIAB. This data encompasses employee-level demographic information, including age, completed education and whether the work was part-time. Data at the establishment level, including the total number of employees, are obtained from the linked Establishment-History-Panel (BHP). A detailed description of LIAB is available in Ruf et al. (2021).

The main analysis focuses on employment spells from 2011-2019. As we will exploit exogenous variation around the cutoff in firm size at 200 employees, we only use observations from firms with between 150 and 250 employees in 2018. For employment spells that did not last an entire year, we keep all observations that include the 30th of June, the date on which the size of firms is recorded. We discard all observations with a zero wage, indicating employment interruptions. This leaves 861,673 relevant observations from 241,372 individuals at 13,330 firms in our main sample. Table 4.1 reports summary statistics of this sample.[8] We observe that workers of the same gender in control firms are comparable to those in the treated firms in terms of age, education and the share of part-time workers.[9]

One limitation of LIAB is the lack of administrative data on hourly wages. Instead, daily wages are calculated based on employer-reported fixed-period wages. The wage data is top-coded for individuals who earn more than the upper earnings limit for statutory pension insurance. In our main analysis, we do not take the censoring into account, but include a robustness check where all censored employment spells are discarded.[10] Although we do not know how many hours an employee worked per week, we do observe whether they worked full-time or part-time. We control for part-time workers in our main regression specifications.

Another limitation of LIAB concerns the fact that the data is limited by the inclusion in the IAB Establishment panel, while administrative data is available for a broader

[8]Source DOI: 10.5164/IAB.FDZD.1906.en.v1, own calculations. We use these data for all results in Section 4.3.

[9]We illustrate the observed gender differences in wages in the raw data in Appendix 4.C.

[10]Censored observations constitute only 1.29 % of our main sample.

	Men		Women	
	Large firms (1)	Small firms (2)	Large firms (3)	Small firms (4)
Daily Wage	94.05	94.32	75.22	72.41
	(50.51)	(52.91)	(45.22)	(43.23)
Age	41.27	41.29	42.72	42.54
	(12.64)	(12.71)	(12.45)	(12.63)
College educated	18.04%	18.08%	18.44%	17.67%
Part-time	15.03%	13.71%	48.61%	49.51%
Firms	4,746	7,743	4,301	6,935
Individuals	59,651	83,663	57,544	60,486
Observations	199,332	285,228	167,662	209,451

Notes: This table reports unconditional means and standard deviations in parentheses of key variables for individuals in large and small firms, split by gender. The descriptive statistics include all data in our panel from 2011 to 2019 in firms with 150 to 250 employees in 2018. 'Age' refers to the employee's age in years, 'College educated' is an indicator of whether the employee has at least some university or university of applied sciences education, and 'Part-time' is an indicator of whether the employee works part time.

Table 4.1: Summary statistics

set of firms. Therefore, we complement our data analysis with a larger data set, as explained in 4.3.5.3. This allows us to obtain even more precise estimates. The downside of this second data set is the time window of observation, as it only includes data up to and including 2018. With the German transparency policy being introduced in January 2018, this second sample only contains one year of post-treatment outcomes. Therefore, we primarily use the smaller LIAB data set.

4.3.3 Identification strategy

We aim to estimate the impact of the wage transparency law on the gender wage gap in affected firms. Our identification strategy relies on the implementation of the wage transparency measure based on the size of the firm. We compare control firms just below the threshold with treated firms just above the threshold, using a difference-in-difference (Diff-in-Diff) analysis.

Equation 4.1 gives the main specification for the Diff-in-Diff approach.

$$\begin{aligned} Y_{ijt} =& \beta_1(Female_i \times Large_j \times Post_t) + \beta_2(Female_i \times Post_t) + \\ & \beta_3(Large_j \times Post_t) + \beta_4(Female_i \times Large_j) + \alpha_i + \\ & \alpha_j + \alpha_t + \delta X_{ijt} + u_{ijt} \end{aligned} \tag{4.1}$$

The outcome Y_{ijt} is the log of the daily wage of individual i, working in firm j in year t. $Female$ is a gender dummy, $Post$ is a dummy indicating whether the observation is from 2018 or 2019 (when the transparency law was active) and $Large$ is a dummy for firms with 200 or more employees in 2018. Note that the right to request wages of comparable workers was only in effect for firms where $Large \times Post$ is equal to one. Throughout the paper, we will use the size of firms, referring to the number of employees observed in 2018 to determine treatment assignment. In a robustness check, we will use the size in the pre-treatment year 2017 instead to avoid any manipulation of size around the cutoff. α_i, α_j and α_t denote individual-, firm- and time-fixed effects. X_{ijt} controls for individual characteristics that vary over time (age squared, education and whether the employee worked part-time).

To study the differential effect of the wage transparency legislation on men and women, we include an interaction between $Female$ and the treated group. We will also report results from gender-specific Diff-in-Diff regressions to evaluate the impact of the policy on male and female wages separately.

β_1 is the coefficient of interest, capturing the change in the gender wage gap in treated firms compared to control firms in the treated period. The main identifying assumption is the parallel-trends assumption. It assumes that the gender wage gap in firms with 200-250 employees evolves over time in the same way as the gap in firms with 150-199 employees (Olden and Møen, 2022). We use an event study to address the plausibility of the parallel-trends assumption in this setting. A difference-in-discontinuity (Diff-in-Disc) approach is used as an additional robustness check, as in Grembi et al. (2016).

4.3.4 Results

Table 4.2 reports the results from our Diff-in-Diff regressions. Overall, we find no effect of the wage transparency law on wages. The first three columns report results of regressions including employee-level time-varying controls. Column (1) gives the results from our main Diff-in-Diff specification. While we confirm that the gender pay gap is reduced in the post-treatment years compared to earlier years (see the coefficient for the $Female \times Post$ interaction; $p < 0.001$), this cannot be attributed to the wage transparency regulation. The coefficient associated with $Female \times Large \times Post$ (β_1 in equation 1) is statistically insignificant, with a point estimate that is indistinguishable from zero ($p = 0.992$). This indicates that the law did not have an effect on the gender pay gap. In other words, the gender wage gap in firms bound by the wage transparency policy did not change in the treated period in a different way than the gender wage gap in the control firms.

Columns (2) and (3) show the impact of the transparency law on male and female wages separately. The coefficients of interest are small and not statistically different from zero ($p = 0.863$ and $p = 0.675$, respectively). We can rule out an impact of more than a 1.5% change in wage for either gender in the 95% confidence intervals. In the joint sample of men and women, we can rule out that overall wages changed by more

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Large × Post	0.0022	0.0009	0.0027	0.0051	0.0044	0.0022
	(0.46)	(0.17)	(0.42)	(0.82)	(0.71)	(0.31)
Female × Large × Post	-0.0001			-0.0028		
	(-0.01)			(-0.36)		
Female × Large	-0.0249			0.0037		
	(-0.83)			(0.17)		
Female × Post	0.0146***			0.0046		
	(3.30)			(0.91)		
Ind. time-varying controls	✓	✓	✓			
Firm FE	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	584,026	325,869	257,544	778,441	435,591	342,066

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.2: Diff-in-Diff estimates of impact of wage transparency law on daily wages

than 1%. The last three columns show that the estimated impact remains close to zero when individual time-varying controls are omitted. Overall, we do not find any evidence of an economically significant impact of the wage transparency regulations on wages.

Next, we will consider whether the law is effective in sub-groups of the German labor force, specifically for employees (not) covered by sectoral bargaining agreement, and whether the regulation resulted in employees seeking alternative employment.

4.3.4.1 The role of collective bargaining agreements

Unions play a prominent role in German industrial relations through bargaining sector-level collective agreements with employer associations and there is evidence that unionization can affect the success of transparency legislation (Cullen and Pakzad-Hurson, 2021). Almost half of all employees in Germany were covered by collective agreements in 2016 (Ellguth and Kohaut, 2019). In our sample of firms with 150 to 250 employees, 70.18% of male and 76.16% of female employees were employed in establishments bound by sectoral or firm-level bargaining agreements in 2018.

An exception for firms bound by a collective bargaining agreement outlined in the transparency law warrants a subgroup analysis by collective bargaining status. If bound by a collective bargaining agreement, it is assumed that workers who perform comparable activities, as defined by being in the same salary scale, receive adequate payments. This so-called 'presumption of adequacy' also applies to firms which use existing sectoral agreements for orientation without being formerly bound by them and implies that the transparency regulation does not allow employees to obtain additional wage information. Furthermore, there is less scope for employees covered by collective bargaining agreements to bargain with their employers individually, as wages and working conditions are set collectively. Even non-union members working for companies subjected to collective wage agreements are generally granted the same benefits. Thus, the transparency law potentially only has an effect in firms that do not adhere to collective bargaining agreements.

We leverage information from the IAB establishment panel to analyze the impact of wage transparency on firms either covered or not covered by a sectoral or firm-level collective bargaining agreement. Using our preferred specification with individual controls, the Diff-in-Diff estimates of interest are not statistically significant, see Figure 4.1. Both in establishments covered by collective bargaining agreement, see Table 4.C.12 in Appendix 4.C.3, and for establishments not covered by collective bargaining agreement, see Table 4.C.13 in Appendix 4.C.3, there is no clear evidence of an effect on wages for men ($p = 0.931$ and $p = 0.940$, respectively) nor women ($p = 0.515$ and $p = 0.344$, respectively)[11]. In other words, there also no significant treatment effects for the sub-sample where we expect the transparency law to be important. These estimates are based on a smaller sample than our main results, as we could only match the collective bargaining status for about half of our main sample.

4.3.4.2 The effect on employment changes

So far, our results demonstrate that wages are not affected by the transparency law. More precisely, we show that the wages in firms with more than 200 employees do not change more after the introduction of the transparency policy compared to wages in firms with fewer than 200 employees. However, the wage transparency regulation may affect workers in other ways. In particular, we investigate whether this regulation impacts the propensity of employees to change their employer. If wage information reveals that an employee's compensation is lower than the comparable other's, the employee might be inclined to search for alternative employment. As employees do not necessarily move to employment in an establishment with a similar number of employees, this would not be captured in our results. Therefore, we consider the

[11]As an additional robustness check, we also include firms that use existing sectoral agreements for orientation in the pool of observations that are affected by collective bargaining, as these firms also benefit from the 'presumption of adequacy'. However, the data coverage for this measure is considerably lower, which implies that this analysis is only based on 69 firms that are not affected by collective bargaining, compared to 162 firms if we do not consider orientation towards sectoral agreements. We again find no consistent evidence in our preferred specification of a differential effect on wages by gender if affected by collective bargaining agreements ($p = 0.142$), or not ($p = 0.974$), see Table 4.C.12 and 4.C.13.

effect of the transparency law on the employee's propensity to switch establishments.

We employ the same Diff-in-Diff regression specification as outlined in Section 4.3.3, Equation 4.1. Instead of using the log of the daily wage as the outcome variable, we define a binary variable that is equal to one if the employee changes within one year the establishment in which they are employed and zero otherwise. We see in our preferred specification that neither male nor female employees are more likely to seek employment at a different establishment due to the transparency regulation ($p = 0.991$ and $p = 0.731$, respectively). See Table 4.C.1 for the regression results.

4.3.5 Robustness checks

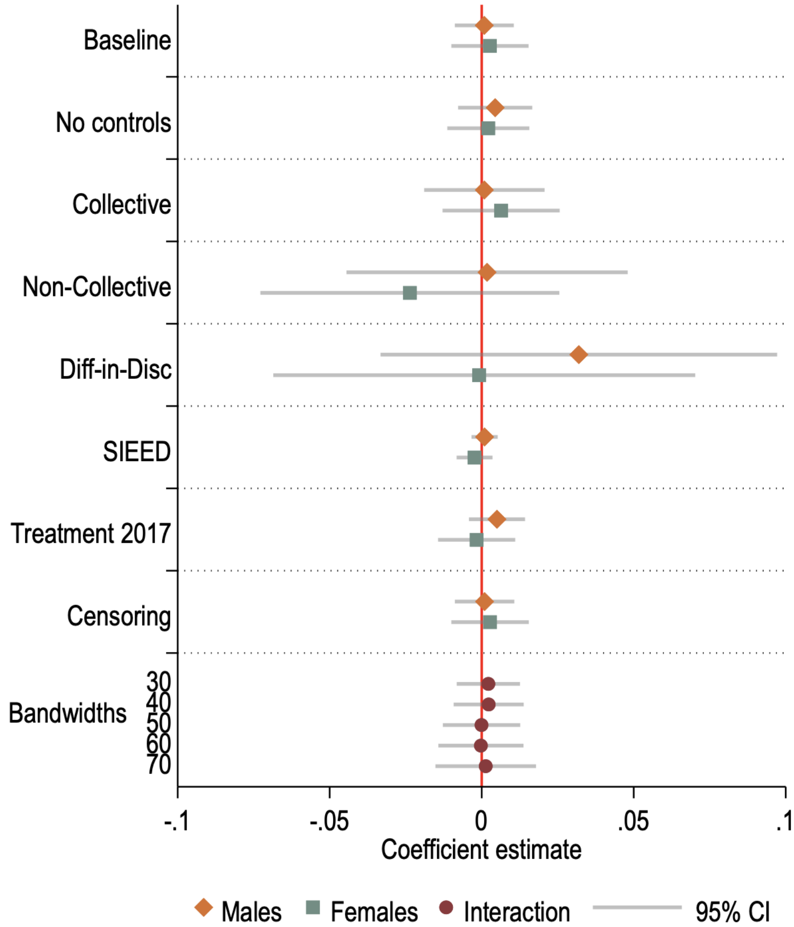
Using a Diff-in-Diff specification, Section 4.3.4 shows that the wage transparency law does not affect wages or the gender pay gap. In this section, we verify that our results are robust and not driven by the details of our specifications. Figure 4.1 provides a first overview of the coefficient estimates of our distinct analyses, demonstrating the robustness of our results. Next, we will lay out the specifics of the robustness checks that we perform.

4.3.5.1 Event study

First, we use an event study specification to evaluate the parallel trends assumption for the Diff-in-Diff specification. We estimate the following model, omitting 2017, the year prior to the reform:

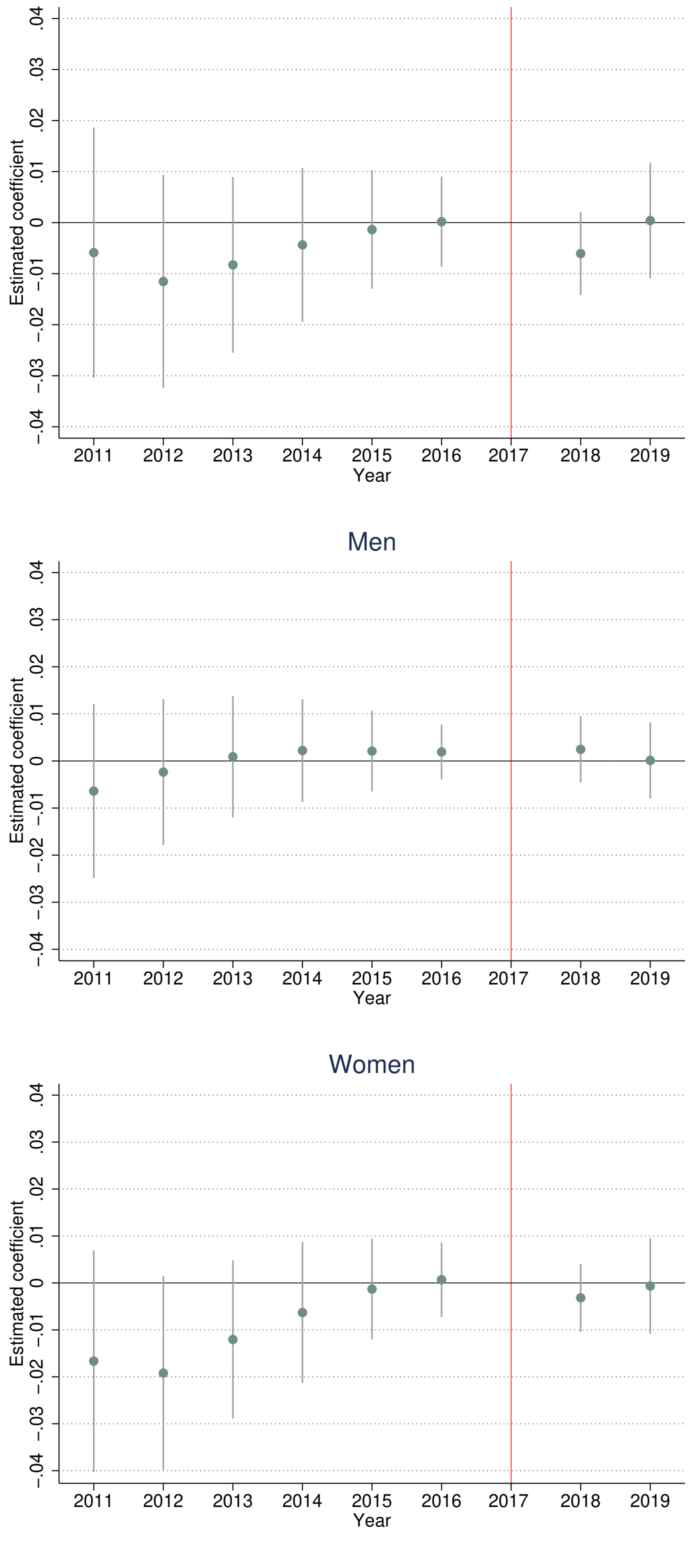
$$\begin{aligned} Y_{ijt} = &\sum_{k=2011}^{2019} \beta_k Female_i \times Large_j \mathbb{1}[t=k] + \sum_{k=2011}^{2019} \gamma_k Large_j \mathbb{1}[t=k] + \\ &\sum_{k=2011}^{2019} \pi_k Female_i + \alpha_t + \delta X_{ijt} + u_{ijt} \end{aligned} \tag{4.2}$$

If there are any pre-policy differences in trends between the treated and control firm, they will be captured by the coefficients β_k in pre-treatment years. The top panel in Figure 4.2 shows the estimated coefficients for β_k. We can see that the estimates are close to zero and do not seem to exhibit a trend in the period between 2011 and 2016, indicating support for the the parallel trends assumption. Furthermore, the estimated coefficients in the post-treatment periods are not statistically significant, suggesting that the transparency policy did not significantly impact the gender wage gap. We can exclude a treatment effect of more than 1.5 percentage points in our 95% confidence interval for both post-treatment years. The bottom two panels in Figure 4.2 display differences in wages in treated and control firms for men and women separately. These again indicate that the reform had no impact on the wages of either gender.



Notes: Coefficient estimates for the robustness checks outlined in Section 4.3.5. ‘Baseline’ refers to the estimates of the Diff-in-Diff regression in columns (2) and (3) of Table 4.2, ‘No controls’ to columns (5) and (6) of Table 4.2. ‘Collective’ gives the Diff-in-Diff estimates when the sample is restricted to employees covered by bargaining agreements in columns (2) and (3) of Table 4.C.12, ‘Non-Collective’ if the sample is restricted to employees not covered by these agreements. in columns (2) and (3) of Table 4.C.13. ‘Diff-in-Disc’ gives the estimates of the Diff-in-Disc analysis presented in columns (2) and (3) of Table 4.C.2. ‘SIEED’ gives the Diff-in-Diff estimates using the SIEED sample presented in columns (2) and (3) of Table 4.C.10. ‘Treatment 2017’ gives the Diff-in-Diff estimates if the number of employees in 2017 is used to determine treatment, see columns (2) and (3) in Table 4.C.3. ‘Censoring’ refers to estimates from the Diff-in-Diff analysis if top-coded observations are discarded, as in columns (2) and (3) in Table 4.C.5. ‘Bandwidths’ refers to the Diff-in-Diff estimates varying the bandwidths left and right of the cutoff, as presented in columns (1) to (5) in Table 4.C.7.

Figure 4.1: Overview of robustness checks



Notes: Event study analysis of the impact of wage transparency regulation on log daily wage. The top figure provides the estimates of the differential impact for women vs. men (β_k in Equation 4.2), the bottom two figures the yearly estimates of $Larger_j$ for separate event study specifications. Firms with more than 200 employees are classified as treated. Individual-, firm- and year-fixed effects are included. Time varying controls include age squared, education and part-time workers. 584,026 observations, including men and women. Error bars indicate the 95% confidence interval. Standard errors are clustered at the firm level.

Figure 4.2: Gender-specific effects of the transparency law

4.3.5.2 Difference-in-discontinuity

An alternative way to address potential biases from differential wage trends for small and large firms is using a Diff-in-Disc estimation introduced by Grembi et al. (2016). This methodology also allows us to control for the impact of any other policy changes at the threshold of 200 employees. In this alternative specification, we consider the following regression:

$$\begin{aligned} Y_{ijt} =& \beta_1 Size_j + Large_j \times (\gamma_0 + \gamma_1 Size_j) + \\ & Post_t[\delta_1 Size_j + Large_j \times (\lambda_0 + \lambda_1 Size_j)] + \alpha_t + \pi X_{ijt} + u_{ijt} \end{aligned} \tag{4.3}$$

$Size_j$ denotes the size of a firm in 2018. λ_0 is the Diff-in-Disc coefficient, which will be estimated separately for men and women. With the Diff-in-Disc estimator, we test whether the discrete jump at the cutoff when approaching from below compared to approaching from above is different for the treatment period compared to control periods. The key identifying assumption for a causal interpretation is the continuity of potential outcomes at the threshold of 200 employees.

Table 4.C.2 gives the results from our main Diff-in-Disc regression. The estimates for gender-specific difference-in-discontinuity coefficients are displayed in columns 2 and 3. The point estimate for the discontinuity in the male sample of 0.032 is statistically insignificant ($p = 0.337$), as is the point estimate for the female sample of 0.001 ($p = 0.981$). This result is also reflected when we interact the Diff-in-Disc estimator with a dummy for women (column (1) in Table 4.C.2), indicating that there are no gender differences in the treatment effect ($p = 0.487$). Overall, these results are qualitatively comparable but less precise than our main Diff-in-Diff specifications. The wage transparency law has no detectable effect on the gender pay gap.

4.3.5.3 Alternative data set

As a further robustness check, we conduct our primary analysis with a different, larger data set. For this, we use the Sample of Integrated Employer-Employee Data (SIEED) by the IAB. SIEED provides administrative data from the same data sources as in our primary analysis. It, however, covers 1.5% of all German establishments, which results in 1,842,584 relevant observations. This is substantially more than in our primary analysis. This larger data set allows us to obtain more precise estimates.

As of 2022, SIEED only includes one post-treatment year. This limits the meaningfulness of the results obtained with this data set, since the initiation of wage negotiations and the accompanying use of wage information might take some time. It is conceivable that we do not observe any impact because the availability of wage information only affects wages in later years. Thus, we do not use the SIEED as our primary sample.

We provide summary statistics and reproduce our results from Section 4.3.4 in Appendix 4.C using SIEED. Both the Diff-in-Diff and Diff-in-Disc results are in line with the findings we presented previously, see Tables 4.C.10 and 4.C.11. The event study in Figure 4.C.3 underlines this. As Figure 4.1 shows, the wage transparency regulation neither significantly affects wages of women ($p = 0.435$) nor men ($p = 0.666$) in 2018. This sample allows us to rule out an effect of more than 1% on the wages of either gender.

4.3.5.4 Alternative regression specifications

We classify whether employees in firms have a right to wage information by the number of employees a firm had in 2018. However, if firms selectively manipulate their size in 2018 around the policy cutoff, the effect estimated in the previous section would be biased. A McCrary test for the continuity of the density of the variable $Size_j$ around the cutoff of 200 employees in 2018 provides no evidence of manipulation ($p = 0.712$). We illustrate the smoothness of the density around the cutoff in Figure 4.C.2. Nevertheless, we use the size of firms in the year prior to the reform as a proxy for treatment to calculate an intention-to-treat effect. Table 4.C.3 in the appendix shows the main outcomes of a Diff-in-Diff analysis using this alternative treatment assignment. The estimates are not significantly different from the main results presented in the last section and do not indicate any treatment effect on male or female wages in our main specification, see also Figure 4.1. Using the same alternative treatment assignment, we show the results of a Diff-in-Disc analysis in Table 4.C.4 in Appendix 4.C. There is again no statistically significant effect.

Wages in our sample are censored, as wages above the upper earnings limit for statutory pension insurance are top-coded. In our main specification, this only affects 1.29% of observations. We address censoring in Appendix 4.C. Here, we discard all top-coded employment spells from our analysis. Independent of the exact specification, we also do not observe a significant impact of the wage transparency regulation if we remove top-coded observations. Table 4.C.5 provides an overview of our Diff-in-Diff analysis on this restricted sample, Table 4.C.6 for our Diff-in-Disc analysis.

Finally, to check whether our conclusions are sensitive to the chosen bandwidth, we provide additional robustness checks with different bandwidths in Appendix 4.C. These confirm our main specification, as Figure 4.1 illustrates. In particular, we include specifications in the range of the optimal bandwidth selected by the data-driven method introduced by Calonico et al. (2020). This does not change our estimates in any meaningful way.

4.4 Experiment

Section 4.3 shows that the German wage transparency law has to date been unsuccessful in reducing the gender pay gap. We now explore potential drivers of this lack of success. Our online laboratory experiment studies the determinants of and potential barriers to a successful wage transparency policy. In this, we focus on how wage transparency can induce changes in beliefs about average wages and the consequences for wage inequality.

First, in Section 4.4.1, we pin down the intuitive arguments in favor of wage transparency as a tool to decrease the gender pay gap and analyze how its effectiveness may depend on the presence of performance information. This theoretical model will provide predictions for the experiment. Next, we outline the experiment designed to test how the endogenous nature of wage information and the environment in which wage information is available impacts the success of wage transparency regulation in Section 4.4.2 and discuss the results in Section 4.4.3.

4.4.1 Theoretical predictions

In this subsection, we examine why and when wage transparency could help decrease the gender wage gap and provide theoretical predictions for our online laboratory experiment. Assume a worker i bargains for a wage w_i with a firm j. In these negotiations, the worker and firm split a pie π between themselves. The worker believes he or she can contribute $\hat{c}_i$ to the firm. The worker further believes that the firm pays comparable workers, that is, workers performing comparable tasks, an average wage of $\hat{\bar{w}}_i$. He or she believes that the average contribution of the comparable workers to the firm is $\hat{\bar{c}}_i$. Consider worker preferences represented by utility $U_i^W(\mathbf{w}, \mathbf{c})$:

$$U_i^W(\mathbf{w}, \mathbf{c}) = w_i - \alpha_i \left(\frac{w_i}{\hat{c}_i} - \frac{\hat{\bar{w}}_i}{\hat{\bar{c}}_i} \right)^2$$

α_i measures a worker's aversion to perceived unfair payment. We define perceived unfair payment as a worker's belief that he or she receives a different piece rate $(w_i/\hat{c}_i)$ than comparable workers $(\hat{\bar{w}}_i/\hat{\bar{c}}_i)$. The worker is therefore not concerned with wage inequalities per se, but holds the meritocratic ideal that the same contribution should result in the same wage.[12] The firm's objective $U_j^F(w_i)$ is to minimize the wage to the

[12]This definition of an unfair wage is in line with the literature on fairness ideals that demonstrates that the source of an inequality matters for its acceptability. Inequalities that are based on merit are more likely to be deemed acceptable, see e.g. Konow (2000), Cappelen et al. (2007) and Almås et al. (2020).

worker:

$$U_j^F(w_i) = \pi - w_i$$

For simplicity, we assume that both worker and firm have an outside option of $d^F = d^W = 0$.

The wage w_i is part of the Nash bargaining solution if it solves the following optimization problem:

$$\begin{aligned}
\max_{w_i} \quad & \left(w_i - \alpha_i \left(\frac{w_i}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)^2 \right) (\pi - w_i) \\
\text{s.t.} \quad & w_i \geq 0 \\
& \pi \geq w_i
\end{aligned}$$

In the absence of information on wages and contributions, a worker's beliefs about his or her contribution and the piece rate of comparable workers, captured by $\hat{c}_i$ and $\hat{\hat{w}}_i/\hat{\hat{c}}_i$, respectively, do not necessarily correspond to the true values, c_i and $\bar{w}/\bar{c}$. Assume that there are two types of workers, a pessimistic and an optimistic type. The first type, type F, has pessimistic beliefs $\hat{c}_i$ about his or her own contribution. The second type, type M, has optimistic beliefs. Type F also has pessimistic beliefs $\hat{\hat{w}}_i$ about the average wages, while M has optimistic beliefs.

Providing information on the true values of c_i and $\bar{w}/\bar{c}$ can shift beliefs. In particular, when receiving information about the true values c_i and $\bar{w}/\bar{c}$, F will update his or her beliefs about c_i and about $\bar{w}/\bar{c}$ positively, type M negatively.

To analyze the impact of belief shifts on wages in the Nash bargaining solution, we first posit that information on the average wage of comparable others only affects beliefs about exactly this average wage of others, $\bar{w}$, and not beliefs about the average contribution $\bar{c}$. Correspondingly, information on the average performance of comparable others does not affect beliefs about the average wage of others. Realize that this is not a trivial assumption. If a worker learns that others earn more than expected, s/he could reasonably infer that this higher pay may be a reward for higher than expected contributions. Unexpectedly high contributions may be perceived as an indication that wages are also higher than expected, to compensate. We will later relax this assumption.

Let w_i^* define the Nash bargaining solution. Inducing a shift in beliefs affects the w_i^*. We show in Appendix 4.A that the Nash bargaining solution has the following properties: w_i^* (1) increases in $\hat{c}_i$, (2) decreases in $\hat{\hat{c}}_i$, and (3) increases in $\hat{\hat{w}}_i$. Intuitively, an increase in $\hat{c}_i$ implies that the own perceived piece rate relative to the comparable workers' decreases, which can be compensated by an increase in w_i. In contrast, if beliefs about comparable workers' average contributions $\hat{\hat{c}}_i$ increase, this entails a decrease in the perceived piece rate of comparable workers. To counteract the perceived inequality in piece rates, w_i needs to decrease. Last, if beliefs about the average wages of others increase, the reverse holds true. The perceived piece rate of comparable workers increases, which a higher w_i can counterbalance.

To derive testable hypotheses from this model, we assume that women are more frequently of the F type, and men more frequently of the M type. As discussed in the introduction, there is some empirical support for this assumption. Men are more confident in their own abilities (see e.g. Niederle and Vesterlund, 2007) and have more optimistic beliefs about average and future wages (Briel et al., 2022). Using this classification, the model permits the following hypotheses, for which we provide the theoretical proofs in Appendix 4.A:

Hypothesis 1. *Providing information about a comparable worker's wage decreases the gender wage gap.*

The change in beliefs $\hat{\bar{w}}_i$ in response to information on $\bar{w}$ will be negative for type M and positive for type F. Since w_i^* increases in $\hat{\bar{w}}_i$, this implies that the wage of women will respond positively to information about a comparable worker's wage, but negatively for men, decreasing the gender wage gap. A similar reasoning leads to the next hypothesis.

Hypothesis 2. *Providing information about a worker's own performance relative to the comparable worker's performance decreases the gender wage gap.*

Given their pessimistic beliefs about their own compared to others' performance, women's beliefs react positively to information on the true value of c_i compared to $\bar{c}_i$. As w_i^* increases in $\hat{c}_i$, revealing the true value c_i compared to $\bar{c}_i$ induces a positive change in the wages of women, at the same time a negative effect is expected for men.

For our next hypothesis, we relax the assumption that information on average wages and contributions of comparable others cannot affect beliefs about average contributions and wages, respectively. Instead, we propose that if the average wage is higher than expected, $\hat{\bar{c}}_i$ will increase. If the average contribution is higher than expected, $\hat{\bar{w}}_i$ will increase. Workers thus expect that there is a positive correlation between the contributions and wages of other workers. For simplicity, we assume that this correlation is the same for types F and M. As a result, the effect of wage information on beliefs about the average piece rate of comparable workers $\bar{w}/\bar{c}$ is now smaller in absolute terms. We will continue to assume that the effect of positive information on $\bar{w}$ as well as negative information on $\bar{c}$ positively affects beliefs about $\bar{w}/\bar{c}$. Intuitively, if a worker learns about higher than expected wages of others, he or she will not decrease beliefs about the average piece rates.

With this more realistic assumption, the arguments brought forward in favor of Hypotheses 1 and 2 are still valid. However, the effects will be less pronounced. In turn, providing information on $\bar{c}$ and $\bar{w}$ simultaneously now distinctively impacts w_i^* in the Nash bargaining solution. Specifically, if both the true values of $\bar{c}$ and $\bar{w}$ are communicated to the worker, there is no adverse effect that reduces the impact on w_i^* of providing this information. Holding the true values $\bar{c}$ and $\bar{w}$ and prior beliefs about these values constant, the effect of providing information on $\bar{c}$ and $\bar{w}$ jointly on $\hat{\hat{w}}_i/\hat{\hat{c}}_i$ is stronger than the aggregate effects of providing information on $\bar{c}$ and $\bar{w}$ separately. As a consequence, given that w_i^* decreases in $\hat{\hat{c}}_i$ and increases in $\hat{\hat{w}}_i$, the effects on w_i^* are stronger when information is provided jointly. This informs our next hypothesis.

Hypothesis 3. *Providing information about a comparable worker's wage and relative performance jointly has a stronger effect on wages than providing this information separately.*

The intuitive implication is that workers cannot use wage information as effectively if they do not know about the corresponding contribution. Higher wages of others can be attributed to higher contributions, which warrant only a smaller increase in the wage of the worker him- or herself to match piece rates.

Our type classification implies that moving to a joint provision of wage and contribution information will benefit women more. This follows from the fact that type F individuals receive on average information that can shift their beliefs $\hat{c}_i$ and $\hat{\hat{w}}_i$ upwards, while it shifts these beliefs downwards for type M. If, however, part of this effect is offset by a change in the respective other belief, this diminishes the differential change in $\hat{\hat{w}}_i/\hat{\hat{c}}_i$ between F and M types. Therefore, F types benefit to a larger extent from joint information provision.

So far, we have interpreted potential differences in the use of information in terms of gender differences. However, we can also utilise the bargaining model to make predictions about the effect of information provision on the wages of type M versus type F using the classification based on beliefs, not gender. In this case, we do not require that the assumptions on male versus female beliefs hold true in our subject pool. Instead, in the analysis, we can classify the subjects based on beliefs and check whether information reduces wage differences between types F and M, irrespective of gender.

4.4.2 Experimental design

Our experiment mimics wage negotiations between a firm and a worker, varying whether and how wage information is provided and whether performance information is given.[13] The experiment consists of two main parts with four periods each. At the start of the experiment, participants are assigned to matching groups of eight.

[13]We pre-registered on the AEA RCT Registry (Brütt and Yuan, 2021).

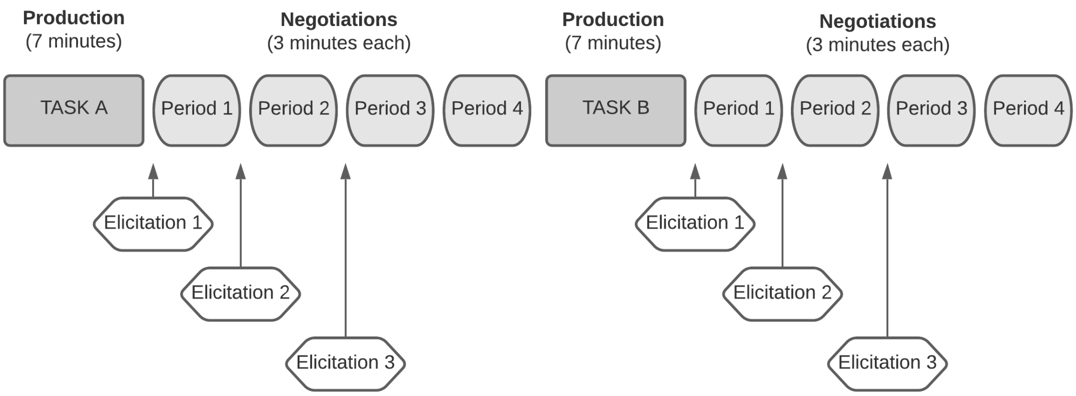


Figure 4.1: Experimental outline

Four are assigned to be a worker, four to be a firm.[14] In each period, one worker is matched with one firm. After each period, subjects are re-matched. We employ a perfect stranger matching within parts. Between parts, the same groups of workers and firms are re-matched.

Figure 4.1 provides an outline of the experimental stages. At the end of the experiment, we elicit risk aversion using the Holt and Laury (2002) multiple price list, and subjects fill in a short questionnaire. We provide the experimental instructions in Appendix 4.E.

4.4.2.1 Production stage

At the start of part 1 and part 2, there is a production stage. In the production stage, workers and firms produce a budget that can be allocated between them in the negotiations. The budget is the sum of the worker's contribution and a fixed firm contribution. Each firm contributes a firm-specific constant to the budget, which is a number drawn from a uniform distribution between 300 and 450 points. This constant is fixed within a part, but re-drawn for each of the two parts.

The worker's contribution is determined in a part-specific production task. The performance in this task determines the worker's contribution to the budget. Workers have to solve as many elements as possible within seven minutes in both tasks. In one part, workers have to produce in the maze task, in the other part in the matrix task.[15] We counter-balance the order of the tasks.

In the maze task, first used in Gneezy et al. (2003), workers have to navigate

[14]Our matching procedure ensures that men and women are distributed as equally as possible to the worker and firms roles within a matching group. The workers were gender balanced. 6 workers did not self-report their gender, or reported "other". We classify the gender of these workers based on the administration data from the laboratories.

[15]While firms cannot produce any output that is added to the budget in the production task, firms also experience the production stage to form an accurate impression of how the worker's contribution is generated.

through mazes on their computer screen. We count the number of mazes they navigate successfully. In the matrix task, introduced by Weber and Schram (2017), workers have to find and then sum up the highest numbers from two matrices with 49 two-digit numbers each. We count the number of correct additions. For each correctly solved element in the production stage, the budget that can be split during negotiations increases by 35 points (for the matrix task) or 20 points (for the maze task).

Both tasks are chosen to be stereotypically male. While studies typically show little evidence for gender differences in the performance in these tasks, spatial reasoning and mathematical skills are often perceived to favor men (Sanchis-Segura et al., 2018).[16] We choose stereotypically male tasks to create an environment where gender differences in wages are likely to emerge from negotiations due to differences in beliefs as described in Section 4.4.1.

4.4.2.2 Negotiation stage

In the first period within a part, all workers enter negotiations[17]. In subsequent periods, workers first unilaterally decide whether they want to enter negotiations. If they do not enter negotiations, workers receive an outside option of 150 points, the remainder is allocated to the firm.[18]

If negotiations occur, workers and firms first submit an initial, non-binding wage proposal. This wage proposal is shown to their negotiation partner during negotiations. Afterward, they enter a three-minute, free-form chat. This stage mirrors the negotiation setup in Exley et al. (2020). Next to the chat, participants can submit and accept wages in a separate field. To agree on a wage, either the worker or the firm has to accept the other side's wage proposal. If the worker and firm agree on a wage, the worker receives this wage and the firm the remainder of the budget. If there is no agreement, both receive zero points.

During the negotiations, only the firm knows the size of the budget that can be split between worker and firm. This allows firms to avoid offering the focal point of an equal budget split. We furthermore do not disclose the exact size of the firm's fixed contribution to the firm or worker. In this way, firms cannot reveal the worker's contribution in the chat in treatments without performance information.

[16] Studies such as Gneezy et al. (2003) and Schram et al. (2019) report no significant gender differences in performance with non-competitive payment and without status ranking, respectively. In an incentivized pre-study run with 100 participants on Prolific, we confirm that these tasks are indeed perceived to favor male participants. See Appendix 4.B for details.

[17] This ensures that the wage we observe of a comparable worker (see Section 4.4.2.3) for all subsequent periods is determined by wage negotiations, not an outside option.

[18] This outside option is set such that even if firms receive the lowest possible draw as their fixed contribution and the worker produces no output, an equitable split would still result in a wage that corresponds to the outside option for the worker. Thus, workers can expect that it is beneficial to enter negotiations.

This negotiation stage is repeated three times after the first period in each part, with re-matching after each period.

4.4.2.3 Treatments

In a 3 × 2 design, we manipulate the information provided during the negotiations along two dimensions, wage information and performance information.

Wage information We vary the provision of wage information between-subjects. Wage information refers to the wage of a 'comparable worker'. We define a worker's comparable worker as the worker who was paired with the current worker's firm in period one. The wage of the comparable worker is comparable in two dimensions. First, as the comparable worker's wage refers to the wage that this worker received in the same part, s/he performed the same task. Second, both workers were paired with the same firm for the wage concerned.[19]

The three between-subject treatments differ in the availability of wage information. The baseline treatments do not provide wage information (*NoWage* treatments), representing the scenario without wage transparency regulation. The second type of treatments provide wage information endogenously (*EndoWage* treatments). Here, workers face the choice of receiving wage information before deciding on whether to enter wage negotiations. Acquiring wage information costs 10 points[20]. The information choice is communicated to the firm. In these treatments, we mimic the wage transparency regulation in Germany, which requires employees to approach their employer in order to acquire wage information. The third type of treatments provide wage information exogenously (*ExoWage* treatments). In contrast to the *EndoWage* treatments, workers here do not face the choice of acquiring wage information. Instead, this is provided for free before the negotiation entry decision. These treatments are closer to a setting where the duty of providing information lies with the employer.

As wage information is created in period one of each part, workers cannot obtain wage information in this period. Treatments, therefore, only differ in periods two to four of each part.

Performance information The treatments *Performance* and *NoPerformance* vary the provision of information about both own performance and the comparable worker's performance. This variation occurs within-subject; participants face the *Performance*

[19]The German wage transparency law mandates that employers provide information about the median comparable worker, while we provide wage information on one worker and not the median wage of all previously matched workers of a firm. We opted to provide the same information in all periods to keep the informational value constant across periods. We can interpret the information that is provided as a signal of the wage of the median worker.

[20]This small but non-negligible cost ensures that we observe whether participants have a strict preference for receiving wage information.

treatment in one part and the *NoPerformance* treatment in the other part. The order of the within-subject treatments and the combination of performance information and working on a specific task are counter-balanced.

4.4.2.4 Belief elicitations

We elicit workers' beliefs about performance and wages at several points during the experiment.[21] First, after each part's production stage, we elicit beliefs about the participant's own performance and the part's comparable worker's performance (*Elicitation 1*). We ask subjects to estimate how many elements were solved correctly. Second, after each part's first negotiation period, we elicit workers' beliefs about the comparable worker's wage (*Elicitation 2*). Third, there are treatment- and choice-contingent elicitations after the second negotiation period in each part (*Elicitation 3*). In treatments and periods without wage information but with performance information, we re-elicit a worker's belief about the comparable worker's wage. Similarly, we re-elicit performance beliefs in treatments and periods without performance information but with wage information.

We elicit beliefs using the binarized scoring rule (Hossain and Okui, 2013). The subjects' estimates are transformed via a quadratic loss function into a probability to win a prize of three Euros.[22] See Appendix 4.E for the detailed procedures and instructions.

4.4.2.5 Experimental procedures

The experiment was conducted online in 24 sessions in May and June 2021, with participants from the subject pools of the CREED laboratory of University of Amsterdam in the Netherlands and the MELESSA laboratory of Ludwig Maximilian University of Munich in Germany. We recruited 528 subjects, 264 each from CREED and MELESSA. We collected observations from 22 matching groups per between-subject variation, eleven from CREED and MELESSA for each between-subject treatment.

Recruiting participants for online experiments from subject pools of university laboratories ensures that participants are aware that practices commonly used at the laboratory, such as no deception, will also apply online. Furthermore, drop-out rates are low even in long experiments.[23] Participants had to correctly answer all comprehen-

[21] Aside from studying belief updating about performance and wages, we can also utilise the elicitations to classify participants into the types described in Section 4.4.1.

[22] In line with recent findings by Danz et al. (2022), we withhold information about the exact incentive structure of the binarized scoring rule to limit biased reporting. Instead, we state that subjects maximize their chance of winning the prize by providing their true beliefs. Subjects can receive more detailed information on the mechanism if they actively request this.

[23] In our experiment, only two participants dropped out after the experiment started. In total, observations from 36 periods had to be discarded from the analysis due to subjects experiencing technical difficulties. This amounts to 2.22% of the data.

sion checks about the experimental instructions before starting the experiment.

At the end of the experiment, point earnings were exchanged for Euro at a rate of one Euro per 25 points. We pay one randomly chosen period from one randomly chosen part, one randomly chosen belief elicitation for the workers, and the risk aversion elicitation. Subjects receive a show-up fee of six Euros and a fee of four Euros for filling out the questionnaire. On average, participants earned 26.59 Euros and the experiment lasted 88 minutes.

4.4.3 Experimental results

As outlined in the pre-analysis plan, we only consider negotiations after the first period, when there is a treatment variation in the available information. In the parametric analysis, we will include controls for the worker's and firm's contributions, and laboratory, period and part fixed effects to test our hypotheses. To account for the dependence of observations within a matching group, we cluster standard errors at the matching group level. When comparing raw means, we will use permutation t-tests (*PmtT-test*).[24]

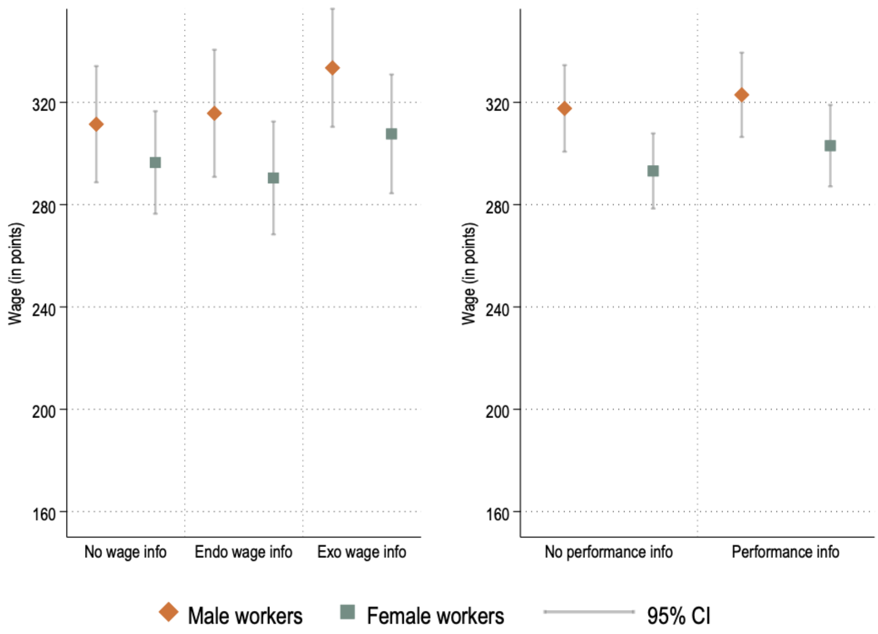
In the following sections, we will first discuss how wages and gender wage differences are affected by wage and performance information, then turn to the effects on negotiation entry. Subsequently, we will take a closer look at the empirical validity of the mechanisms suggested in Section 4.4.1.

4.4.3.1 The effect of transparency on wages

Figure 4.2 provides an overview of average wages by gender in each wage-information treatment. Note that the worker's wage is equal to the outside option of 150 points if he or she did not enter negotiations. If the worker entered negotiations, it is equal to the agreed upon wage, minus the incurred costs of wage information in treatments with endogenous wage information.

The effects of wage information The left panel of Figure 4.2 illustrates wage differences by gender and wage-information treatment. Table 4.1 presents regression results. Including the described control variables and fixed effects, we employ a linear regression of the worker's wage on dummy variables for the wage information treatments (*EndoWage* and *ExoWage*), columns (1) and (2), an indicator for the worker being female, column (3), and the fully interacted variables of the worker's gender and treatment indicators, column (4).

[24] Here, we will average observations on an individual level or, for the comparison of wage differences, on a matching-group level. Permutation t-tests are more powerful than traditional t-tests (Moir, 1998; Schram et al., 2019).



Notes: Comparison of mean wages by gender, varying wage information (left) and performance information (right). Error bars indicate the 95% confidence interval. Standard errors are clustered at the matching-group level.

Figure 4.2: Treatment comparison of gender differences in wages

We confirm in the laboratory that overall, wages are not significantly affected by the introduction of a wage transparency policy that requires workers to ask for wage information ($p = 0.643$; regression (1) in Table 4.1). On average, workers earn a wage of 303.98 points in *NoWage* and 302.95 points in *EndoWage* (*PmtT-test*; $p = 0.9159$). Overall, workers pay for wage information in 47.57% of the decisions. 83.91% of workers request wage information at least once. This documents a substantial demand for wage information if the associated monetary costs are low. Nevertheless, workers do not benefit from the introduction of the type of wage transparency policy that resembles the law discussed in Section 4.3.

Compared to these two treatments, the introduction of exogenous wage information in *ExoWage* has a positive, albeit only marginally significant, effect on wages ($p = 0.076$; regression (2) in Table 4.1). We estimate that exogenously provided wage information increases the workers' wages by 14.65 points. In *ExoWage*, workers earn 320.78 points, 6% more than in the other two wage-information treatments (*PmtT-test*; $p = 0.067$).

	Worker's wage							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Worker contribution	0.55***	0.55***	0.55***	0.55***	0.55***	0.55***	0.55***	0.55***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Firm contribution	0.24**	0.24**	0.23**	0.24**	0.24**	0.24**	0.25**	0.25**
	(0.09)	(0.09)	(0.09)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)
Endo wage	-4.92			-5.44			-1.40	-5.02
	(10.57)			(14.05)			(13.08)	(17.74)
Exo wage	12.18	14.65*		14.55			15.33	22.13
	(9.13)	(8.13)		(12.48)			(11.96)	(16.09)
Female			1.99	3.38		4.74		7.09
			(5.82)	(12.34)		(8.98)		(18.05)
Endo wage × Female				0.95				7.00
				(14.52)				(22.17)
Exo wage × Female				-4.80				-13.88
				(15.67)				(22.49)
Performance					10.73**	13.45*	15.25*	19.05
					(5.08)	(7.57)	(8.63)	(13.25)
Performance × Female						-5.43		-7.67
						(10.13)		(20.44)
Performance × Endo wage							-7.14	-1.19
							(12.34)	(18.45)
Performance × Exo wage							-6.32	-15.39
							(12.58)	(18.59)
Performance × Endo wage × Female								-11.70
								(25.31)
Performance × Exo wage × Female								18.53
								(25.94)
Constant	-0.44	-3.27	2.88	-2.92	-2.96	-6.27	-9.80	-14.16
	(35.10)	(34.11)	(35.77)	(37.43)	(35.92)	(37.10)	(34.78)	(38.10)
Part FE	✓	✓	✓	✓	✓	✓	✓	✓
Period FE	✓	✓	✓	✓	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1548	1548	1548	1548	1548	1548	1548	1548
Clusters	66	66	66	66	66	66	66	66
R-squared	0.265	0.265	0.262	0.265	0.264	0.264	0.267	0.268

Notes: Results are from ordinary least squares regression of the worker's wage. Worker contribution is a control for the worker's contribution to the negotiation pie, Firm contribution for the firm's contribution to the negotiation pie. Endo wage and Exo wage are indicators of whether wage information was provided endogenously or exogenously, respectively. Female indicates whether a participant is female. Performance is an indicator of whether information of the workers' performances is provided. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.1: The effect of wage and performance information on wages

This suggests that the accessibility of wage information indeed matters. Note that we only implement a small cost of 10 points for obtaining this information in *EndoWage*. Yet, this treatment shows virtually identical outcomes for workers compared to *NoWage*. This is in line with the notion that providing this information only on request is a barrier to the utilization of wage information. Possible reasons are fear of backlash or wrong perceptions about its usefulness, which may limit uptake. In Section 4.4.3.3, we will discuss this second potential reason. Workers seem to take advantage of wage information only when it is provided exogenously.

Next, we consider gender-specific effects. Male workers earn on average a wage of 320.27 points in our experiment, female workers significantly less at 298.11 points (*PmtT-test*; $p = 0.029$). In regression (3) of Table 4.1, we see that this gap in our experiment disappears if we control for the worker's and firm's contribution ($p = 0.733$).[25] We can nevertheless study whether the treatments have a differential impact on male and female workers. In particular, our experiment provides a setting where women, on average, are paired with a comparable worker who obtained a wage that is 16.78 points higher than their own wage. In comparison, men face comparable workers who obtained a wage that is 16.38 points lower than the average male worker's wage. This difference is significant (*PmtT-test*; $p = 0.018$). Gender differences in wages induce gender differences in the information that is provided. Furthermore, women are significantly more pessimistic about the wage of the comparable worker than men (*PmtT-test*; $p = 0.067$). Therefore, wage information has the potential to shift women's beliefs to a larger extent.

As in the field, a wage transparency policy that requires workers to ask for wage information themselves (in *EndoWage*) does not have a differential effect on male and female workers. It does not reduce the unconditional gender pay gap compared to the *NoWage* treatment without any wage information ($p = 0.948$; regression (4) in Table 4.1). In *NoWage*, male workers earn 5% more than female workers, in *EndoWage* 9% more. These wage gaps are statistically indistinguishable (*PmtT-test*; $p = 0.713$).

Our results so far confirm the findings from the field. As a next step, we want to study whether removing the barrier to wage information alleviates its ineffectiveness in our setting. We do not find any support for this. The unconditional gender wage gap in treatment *ExoWage* amounts to 8%, which is no reduction compared to the wage gap in *NoWage* (*PmtT-test*; $p = 0.671$) and similar to the gap in *EndoWage* (*PmtT-test*; $p = 0.976$). Together, our results provide evidence that in our context a move to more accessible wage information does increase overall wages, but this effect is not gender specific ($p = 0.760$; regression (4) in Table 4.1). Therefore, the accessibility of wage information on its own does not lead to a reduction in the gender pay gap.

Since our experimental design isolates the effect that wage transparency regula-

[25] The worker's contribution was about 10% lower for female workers in both the maze task (*PmtT-test*; $p = 0.005$) and the matrix task (*PmtT-test*; $p < 0.001$).

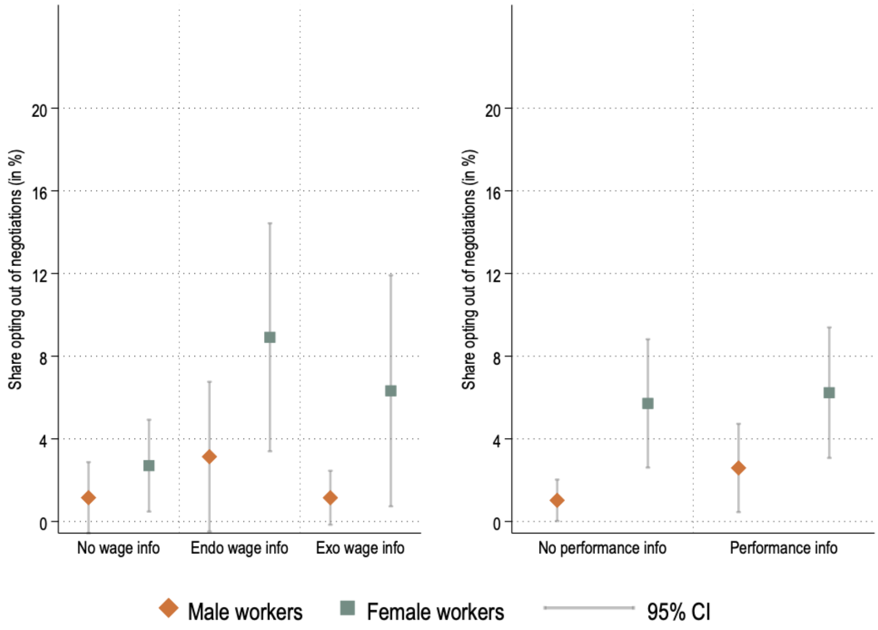
tion has by changing beliefs about wages and the role of these beliefs in bargaining, we can conclude that there is no evidence in favor of this channel leading to a reduction in the gender pay gap. This holds irrespective of how wage information is provided. We therefore cannot reject the null hypothesis of no effect of wage information on the gender wage gap in favor of Hypothesis 1. Section 4.4.3.4 discusses the interaction of beliefs and information provision in more detail.

The effects of performance information The right panel of Figure 4.2 depicts wages by gender and performance-information treatment. In Table 4.1, we provide results of a linear regression of the worker's wage on a dummy variable for the performance information treatment (*Performance*), column (5), and the fully interacted variables of the indicator *Performance* with the indicator of the worker being female, column (6).

Overall, the workers' wages are slightly higher if they know their performance and the comparable worker's performance. Workers earn 312.99 points with performance information compared to 305.45 points without this information. While this difference is small, it yields a significant effect of performance information on wages in our parametric specification ($p = 0.039$; regression (5) in Table 4.1). We estimate that providing performance information increases the workers' wage by 10.73 points. This suggests that a worker's bargaining power increases if the informational asymmetry between worker and firm is reduced. We also observe that workers receive wages that better reflect their performance in treatment *Performance*. In the presence of performance information, workers receive 0.14 points more for every point they have contributed to the negotiation budget ($p = 0.003$; regression (2) in Table 4.D.1).

As observed for wage information, performance information also has no significant effect on the gender pay gap in our experiment ($p = 0.593$; regression (6) in Table 4.1). Female workers do not exploit their knowledge of their relative performance in negotiations more than male workers do or vice versa. Therefore, we cannot reject a null effect of performance information on the gender wage gap in favor of Hypothesis 2.

The effects of a joint provision of wage and performance information We concluded that exogenously providing wage and performance information both have a small, but significant effect on wages. Now, we want to study whether the joint provision of these two types of information can enhance their effectiveness. This would point to wage transparency regulation working better in environments where performance is easily observable. In Table 4.1, we provide the fully interacted model including indicators of the treatments *Performance* with *EndoWage* and *ExoWage*, column (7), also including interactions with the indicator of whether the worker is female



Notes: Comparison of the share of workers opting out of negotiations by gender, varying wage information (left) and performance information (right). Error bars indicate the 95% confidence interval. Standard errors are clustered at the matching-group level.

Figure 4.3: Treatment comparison of gender differences in negotiation opt-outs

in column (8).

We do not find a meaningful interaction effect of performance and wage information. The effect of providing performance information is not significantly different with wage information compared to without wage information ($p = 0.565$ for *EndoWage* and $p = 0.617$ for *ExoWage*); regression (7) in Table 4.1). Moreover, there is no distinct interaction effect of joint information provision for women compared to men ($p = 0.646$ and $p = 0.478$, respectively; regression (8) in Table 4.1). Hence, we also cannot reject the null hypothesis of no effect of joint information provision in favor of Hypothesis 3.

4.4.3.2 The effect of transparency on negotiation entry

In this section, we analyze whether the availability of information affects the workers' willingness to negotiate. Considering negotiation entry is essential, as not entering negotiations usually entails negative payoff consequences. Controlling for differences in contributions by workers who do and do not select into negotiations, opting out of negotiations reduces the worker's wage by 101.01 points ($p < 0.001$). Figure 4.3 depicts the share of male and female workers opting out of negotiations in each treatment.

Whereas performance information does not significantly impact the worker's propensity to enter negotiations (*PmtT-test*; $p = 0.490$), wage transparency deters workers from entering negotiations. Compared to the *NoWage* treatment, significantly more workers opt out of negotiations in *EndoWage* and *ExoWage* (*PmtT-test*; $p = 0.068$, pooling observations from *EndoWage* and *ExoWage*). Importantly, this effect is gender specific. Our experiment replicates the common finding in the literature that women opt out of negotiations significantly more often. Female workers opt out of 6% of all negotiations, male workers only out of 2% of negotiations (*PmtT-test*; $p = 0.004$). This gender difference is primarily driven by women's response to wage transparency. Without wage information, there are no gender differences in the willingness to enter negotiations (*PmtT-test*; $p = 0.379$), but differences emerge in the information treatments (*PmtT-test*; $p = 0.009$, pooling observations from *EndoWage* and *ExoWage*).[26]

The results are in line with women avoiding the social comparison that negotiations with wage information entail. Wage information reveals crucial information on the worker's social status and ranking, which may result in women opting out of negotiations more often due to gender differences in status-ranking aversion (Brandts et al., 2020) and different responses to public self-assessments (Haeckl, 2022).

Note that in our experiment, the small share of decisions to opt out of negotiations means that the gender difference in entry decisions in *ExoWage* does not imply that the gender wage gap increases under wage transparency.[27] If women are more likely to forego the benefits from negotiations if wage information is freely available, this nevertheless results in substantial wage losses for these workers.

4.4.3.3 Endogenous wage information

Next, we turn to potential barriers to the usefulness of endogenous wage information. If requesting wage information is not beneficial for workers, wage policy that requires workers to ask for the information might fail. So we now focus on who is requesting and who is benefiting from wage information.[28] Table 4.2 presents regression results restricting the sample to observations from *EndoWage*. We regress the binary choice of requesting wage information on the worker's contribution and the usual set of controls in column (1), add an indicator for female in column (2), and split the sample by gender in columns (3) and (4).

Overall, women request wage information about five percentage points less often

[26]We observe this gender difference both in *EndoWage* (*PmtT-test*; $p = 0.094$), and *ExoWage* (*PmtT-test*; $p = 0.046$). The regression results for this subsection can be found in Table 4.D.2. It provides the results of OLS regressions of the participant's choice to opt out of negotiations on a gender dummy, the treatment indicators *Wage* and *Performance*, as well as their interactions.

[27]We show in Appendix 4.D, Table 4.D.3, that the wage gap is not affected by treatments *EndoWage* or *ExoWage* if we only consider workers who enter negotiations. Therefore, the fact that gender differences in entry decisions under wage transparency do not result in an increase of the gender pay gap is not the result of a change in wages by women entering negotiations. These women do not benefit from wage information and thus do not compensate for the loss incurred by women who opt out. Instead, the share of choices to opt out of negotiations is too low to significantly affect the gender pay gap.

[28]This section is of an exploratory nature and was not pre-registered.

	Requested wage information			
	(1)	(2)	(3)	(4)
Worker contribution	0.04**	0.04*	0.08*	-0.00
	(0.02)	(0.02)	(0.05)	(0.03)
Female		-0.05		
		(0.08)		
Constant	0.47***	0.51***	0.30	0.65***
	(0.10)	(0.12)	(0.19)	(0.12)
Part FE	✓	✓	✓	✓
Period FE	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓
Sample	*EndoWage*	*EndoWage*	*EndoWage Men*	*EndoWage Women*
Observations	515	515	255	260
Clusters	22	22	22	22
R-squared	0.066	0.068	0.079	0.075

Notes: Results are from ordinary least squares regression of the worker's binary decision to request wage information. Worker contribution is a control for the worker's contribution to the negotiation pie (in hundred units), Female indicates whether a participant is female. Standard errors are clustered at the matching-group level and shown in parentheses. Sample refers to the treatment(s) from which the observations for the analysis stem
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.2: The determinants of requesting wage information

than men, a difference that is statistically insignificant ($p = 0.540$; regression (2) in Table 4.2). More productive workers, on the other hand, are significantly more likely to ask for wage information ($p = 0.050$; regression (1) in Table 4.2). Endogenous wage transparency policies, therefore, are more likely to have an impact on the negotiations of high-performing individuals. Interestingly, this effect is entirely driven by the behavior of male workers ($p = 0.079$ for men $p = 0.954$ for women, regression (3) and (4) in Table 4.2). This effect could be another consequence of (high-performing) men being more inclined to seek social comparisons, as discussed in Section 4.4.3.2.

The rest of this subsection analyzes how the choice of requesting wage information affects wages. Endogenous wage transparency policies are only effective if individuals who request wage information actually benefit from this request. Table 4.3 gives the results of a linear regression with the previously outlined controls and fixed effects of the worker's wage on an indicator of the worker requesting wage information in treatment *EndoWage*, column (1), including an interaction of this choice with the worker's contribution in column (2). In columns (3) and (4), the analysis is split by gender. Column (5) only includes observations from *ExoWage* and observations from individuals choosing wage information in *EndoWage*, regressing the worker's wage on an indicator for treatment *ExoWage* and the interaction of *ExoWage* with the worker's contribution.

We first test the effect of requesting wage information in the *EndoWage* treatment. There is no significant effect of requesting wage information on negotiated wages ($p = 0.144$; regression (1) in Table 4.3). If anything, the effect is more likely to be

	Worker's wage				
	(1)	(2)	(3)	(4)	(5)
Worker contribution	0.57***	0.70***	0.78***	0.64***	0.50***
	(0.03)	(0.05)	(0.05)	(0.08)	(0.03)
Firm contribution	0.06	0.07	0.11	-0.01	0.25**
	(0.18)	(0.18)	(0.21)	(0.17)	(0.12)
Info choice	-17.77	76.86**	76.07*	77.56*	
	(11.72)	(30.23)	(40.55)	(43.42)	
Info choice × Worker contribution		-0.25***	-0.26**	-0.25*	
		(0.08)	(0.10)	(0.14)	
Exo wage					26.14**
					(11.69)
Constant	67.34	19.77	-4.00	45.13	4.63
	(66.20)	(69.78)	(80.52)	(72.36)	(42.52)
Part FE	✓	✓	✓	✓	✓
Period FE	✓	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓	✓
Sample	*EndoWage*	*EndoWage*	*EndoWage Men*	*EndoWage Women*	*WageInfo*
Observations	515	515	255	260	759
Clusters	22	22	22	22	44
R-squared	0.272	0.284	0.294	0.307	0.240

Notes: Results are from ordinary least squares regression of the worker's wage. Worker contribution is a control for the worker's contribution to the negotiation pie, Info choice indicates whether worker requested wage information in *EndoWage*, ExoWage is an indicator for *ExoWage*. Standard errors are clustered at the matching-group level and shown in parentheses. Sample refers to the treatment(s) from which the observations for the analysis stem, *WageInfo* refers to observations from *ExoWage* and individuals choosing wage information in *EndoWage*.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.3: The effect of requesting wage information on wages

negative, with a point estimate of -17.77 points.

Although there is no overall effect of requesting wage information, this pooled analysis hides an important heterogeneity. Requesting wage information helps low performers and hurts high performers ($p = 0.006$; regression (2) in Table 4.3). This is the case both for male and female workers ($p = 0.015$ and $p = 0.096$; regression (3) and regression (4) in Table 4.3, respectively). Intuitively, the wage information provides an anchor for the negotiations, which is, on average, comparatively low for high-performing workers. Without wage information, highly productive workers earn more on average, and the comparable wage is likely to be lower than the wage they would receive without wage information. The reverse is true for low performers. Therefore, the anchor is favorable for low-performing individuals only. As we have previously seen that high-performing workers are more likely to request this information, endogenous wage transparency policies might then fail to improve overall wages.

Finally, we compare the wages of workers who request wage information in *EndoWage* to those that receive it exogenously in *ExoWage*. Here, both comparison groups acquire (endogenously or exogenously) wage information. Nevertheless, wages differ.

We observe significantly higher wages (by 11%) in *EndoWage*, where the information acquisition does not result from an active choice (*PmtT-test*; $p = 0.018$). However, after controlling for the higher performance of those who request wage information, we estimate that the choice to acquire wage information (compared to the exogenous provision) reduces wages by 26.14 points ($p = 0.031$; regression (5) in Table 4.3).[29] This result hints that endogenous wage information will not only reach fewer workers (due to limited take-up), but the workers who do request the information may also benefit less from it than workers if wage information is provided exogenously. This therefore provides further evidence suggesting that endogenous wage transparency may not be optimal.

4.4.3.4 The role of beliefs

Our experiment addresses whether wage information can reduce wage inequality by correcting beliefs about others' wages and relative performance. We now zoom in on this mechanism. For this, we take a closer look at the effects of these types of information on beliefs and the role that incorrect beliefs play in determining negotiation outcomes.

Type classification Previously, we established that controlling for a worker's contribution reduces the gender wage gap in our experiment. This, however, does not necessarily mirror actual labor markets. To study whether belief changes can provide a channel through which wage transparency can be an effective tool, we now directly classify individuals based on their beliefs, not their gender.[30] Do pessimistic individuals benefit more from learning about others' wages and underconfident individuals more from performance information?

Following the theoretical analysis in Section 4.4.1, we utilize two types of beliefs for our classification. First, we use a subject's beliefs about the comparable worker's wage (from *Elicitation 2*). Subjects with beliefs about the comparable worker's wage that exceed the actual wage of the period's comparable worker are classified as 'Optimistic'. Second, we use a subject's belief about performance in the production task (from *Elicitation 1*). As information about own performance and the period's compara-

[29] Any difference in the wages of these two groups reflects the costs of 10 points for acquiring wage information and could be driven by selection in *EndoWage*. The workers who choose wage information are a non-random subsample of the pool of workers. For instance, high-performing individuals are more likely to request wage information. Furthermore, it is possible that workers with low negotiation skills are more likely to request wage information, and that they would have received lower wages regardless of information provision. There is, however, some suggestive evidence that selection is not the main driver of this effect. We show in Table 4.D.4 that workers who do not request wage information in *EndoWage* receive comparable wages as workers in *NoWage*, so these two samples of individuals reach similar outcomes. This suggests that the interaction of receiving wage information and choosing to acquire wage information is crucial.

[30] We pre-registered this approach.

ble worker's performance are always provided jointly, we classify subjects depending on whether they were 'Overconfident' in their relative performance.[31]

Belief updating After workers receive wage or performance information, we re-elicit their beliefs, as explained in Section 4.4.2. Now, we compare wage elicitations from *Elicitation 2* and *Elicitation 3*. This allows us to investigate whether information about wages informs beliefs about performance and vice versa. Indeed, beliefs about the comparable worker's performance are affected by wage information. If workers receive wage information, in *EndoWage* or in *ExoWage*, they update their beliefs more negatively about the comparable worker's performance the more they overestimated the comparable worker's wage ($p = 0.033$; regression (2) in Table 4.D.5). Thus, surprisingly low wages are partially attributed to lower-than-expected performance. Similarly, individuals that were too optimistic about the comparable worker's performance update their beliefs more negatively about the comparable worker's wage if performance information is provided ($p = 0.007$; regression (4) in Table 4.D.5). Therefore, it is important to consider the observability of performance when wage transparency is implemented. See Table 4.D.5 for the regression analysis.

The effects wage and performance information on negotiation outcomes We document the results of linear regressions of the worker's initial wage request (Table 4.4) and the worker's wage (Table 4.5) on the worker's type in columns (1) and (3), including interactions of the worker's type and treatment in columns (2) and (4), with the usual controls and fixed effects. For both tables, we use the full sample in columns (2) and (4) and restrict the sample to individuals in treatment *NoWage* in column (1) and individuals in treatment *NoPerformance* in column (3).

As a first test of whether the type classification predicts negotiation behavior in the hypothesized way, we analyze the effect of information on the workers' initial wage requests in negotiations. Studying initial wage requests allows us to see the different types' responses to information when this has not yet been affected by the firm's behavior or the negotiations in the chat.

The classification of 'Overconfident' and 'Optimistic' workers predicts initial wage requests in the hypothesized way, even after controlling for the worker's contributions. Both optimistic workers ($p = 0.004$, regression (1) in Table 4.4) and overconfident

[31] In other words, individuals classified as 'Overconfident' believe that the difference between their performance and the period's comparable worker's performance exceeds the actual difference. This definition is similar to the overplacement definition of overconfidence found in the literature (Moore and Healy, 2008), although it refers to overestimation of the relative number of questions solved, rather than overestimation of the relative rank. Outliers, with beliefs exceeding 60 correct answers, are excluded from this analysis. These participants likely reported their beliefs about worker contribution, rather than the number of correct answers, and constitute only 1% of observations. This does not affect the results of our analysis in any meaningful way.

	Worker's initial offer			
	(1)	(2)	(3)	(4)
Worker contribution	0.51***	0.52***	0.50***	0.55***
	(0.08)	(0.04)	(0.05)	(0.04)
Optimistic	40.38***	39.93***		
	(12.29)	(12.25)		
Endo wage		9.71		
		(13.22)		
Exo wage		11.24		
		(12.62)		
Endo wage × Optimisticic		-6.17		
		(14.77)		
Exo wage × Optimisticic		-38.39***		
		(14.05)		
Overconfident			28.77***	32.74***
			(7.61)	(7.29)
Performance info				26.81***
				(9.13)
Performance info× Overconfident				-45.70***
				(9.91)
Constant	202.03***	204.81***	215.62***	195.33***
	(28.49)	(18.42)	(21.86)	(18.02)
Part FE	✓	✓	✓	✓
Period FE	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓
Observations	509	1486	743	1469
Clusters	22	66	66	66
R-squared	0.278	0.275	0.248	0.268

Notes: Results are from ordinary least squares regression of the worker's initial request. Worker contribution is a control for the worker's contribution to the negotiation pie. Endo wage and Exo wage are indicators of whether wage information was provided endogenously or exogenously, respectively. Performance is an indicator of whether information of the workers' performances is provided. Optimist indicates that a subject's beliefs about the comparable worker's wage are too optimistic, Overconfident indicates that a subject's beliefs about his or her own performance relative to the comparable worker's are too optimistic. Standard errors are clustered at the matching-group level and shown in parentheses.

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

Table 4.4: The type-specific effect of performance and wage information on initial wage requests

workers ($p < 0.001$; regression (3) in Table 4.4) request significantly higher wages in the absence of the relevant information, in line with our model's predictions.[32]

However, when information is provided to correct these misspecified beliefs, initial wage requests change in the expected direction: compared to the other type, optimistic workers reduce their initial demand by 38 points in *ExoWage* ($p = 0.008$; regression (2) in Table 4.4) and overconfident workers reduce their demand by 46 points in *Performance* ($p < 0.001$; regression (4) in Table 4.4). This gives a first indication of the potential power of information: After correcting beliefs, the initial wage requests of optimistic and overconfident individuals are no longer higher than the demands by other types. However, at this point, it is not clear whether this effect translates into a change in negotiated wages. Therefore, we will now study whether realized wages are affected in a similar way.

We first consider heterogeneous treatment effects depending on whether beliefs about the wage of the comparable worker are too optimistic. Individuals with too optimistic beliefs earn significantly less in the absence of wage information ($p = 0.016$; regression (1) in Table 4.5). This is potentially driven by optimists negotiating for unrealistically high wages, which leads to a negotiation breakdown. In line with this, optimists are significantly more likely to face a breakdown of negotiations, where workers and firms fail to agree and both receive a payoff of zero ($p < 0.001$; regression (1) in Table 4.D.6). However, there is no evidence that wage information improves the outcomes for Optimists ($p = 0.997$ for *Endo Wage*, $p = 0.463$ for *Exo Wage*; regression (2) in Table 4.5). Thus, wage information only changes initial asks by overconfident individuals, without affecting the ultimate negotiation outcomes. Correcting beliefs about wages, therefore, only has an intermediate effect on those individuals in our sample that could benefit from this information.

Next, we consider the effect of performance information depending on whether an individual is overconfident or underconfident. The wages of underconfident individuals increase if performance information is provided ($p = 0.017$; regression (4) in Table 4.5). We estimate that underconfident individuals increase their wages by 21 points, whereas overconfident individuals are not affected by performance information (point estimate of $21.06 - 18.07 = 2.99$ points, $p = 0.096$ for the interaction effect). In line with our theoretical predictions, this suggests that underconfident individuals gain from performance information that corrects their pessimistic beliefs. In contrast to the effect of wage information, the correction of beliefs about relative performance is thus also powerful in affecting wages, not only intermediate outcomes such as initial wage requests.

[32]Note that this initial proposal was made before the unstructured negotiations started, but after the provision of wage and/or performance information.

	Worker's wage			
	(1)	(2)	(3)	(4)
Worker contribution	0.58***	0.58***	0.49***	0.55***
	(0.04)	(0.02)	(0.04)	(0.03)
Firm contribution	0.28	0.22**	0.33**	0.25**
	(0.17)	(0.09)	(0.13)	(0.10)
Optimistic	-33.19**	-33.78***		
	(12.71)	(12.28)		
Endo wage		-4.52		
		(11.87)		
Exo wage		7.53		
		(10.35)		
Endo wage × Optimistic		-0.06		
		(16.14)		
Exo wage × Optimistic		12.13		
		(16.45)		
Overconfident			5.70	10.55
			(9.72)	(9.61)
Performance info				21.06**
				(8.60)
Performance info× Overconfident				-18.07*
				(10.70)
Constant	-12.48	13.68	-18.05	-13.23
	(63.31)	(35.35)	(48.90)	(38.02)
Part FE	✓	✓	✓	✓
Period FE	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓
Observations	519	1548	770	1530
Clusters	22	66	66	66
R-squared	0.292	0.278	0.247	0.263

Notes: Results are from ordinary least squares regression of the worker's wage. Worker contribution is a control for the worker's contribution to the negotiation pie, Firm contribution for the firm's contribution to the negotiation pie. Endo wage and Exo wage are indicators of whether wage information was provided endogenously or exogenously, respectively. Performance is an indicator of whether information of the workers' performances is provided. Optimistic indicates that a subject's beliefs about the comparable worker's wage are too optimistic, Overconfident indicates that a subject's beliefs about his or her own performance relative to the comparable worker's are too optimistic. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.5: The type-specific effect of performance and wage information on wages

4.5 Conclusion

Wage transparency regulation has become an increasingly popular policy tool. Studies on the diverse wage transparency policy landscape can guide the design of future regulations. This is particularly relevant in light of efforts by the EU to establish wage transparency standards. Ours is the first study to look into a unique wage transparency law introduced by Germany, where employees are given the right to request wage information. Using plausibly exogenous variation in whether firms have to comply with this regulation, we do not find any impact on wages or the gender pay gap.

In an online laboratory experiment, we examine several mechanisms underlying the policy's ineffectiveness that can inform future policies. We address the way in which wage information is currently provided, with employees needing to actively request this. If wage information is provided exogenously instead of endogenously, we see that wages increase. This suggests an increase in the workers' bargaining power if wage information is provided by default. In part, the ineffectiveness of endogenously compared to exogenously provided wage information is driven by workers requesting wage information who do not effectively utilize this information. Crucially, the gender wage gap, however, is also not affected by exogenously provided wage information. Moreover, female workers enter negotiations less often if wage information is provided exogenously, suggesting that wage transparency may also backfire.

As a complimentary transparency measure, we study performance information. Performance information increases workers' wages, but does not affect the gender pay gap. Our study underlines why it is nevertheless important to consider performance information when designing transparency regulations. When performance comparisons are difficult, the effect of wage transparency on correcting beliefs about a worker's fair compensation may be dampened. Individuals could attribute the news they receive about others' wages to performance differences instead of only updating their beliefs about wages.

Our research is a first step that indicates that 'pay information rights' do not perform as well as previously studied 'pay information duties', such as investigated by Duchini et al. (2020) and Bennedsen et al. (2022). As a next step, the effect of wage transparency regulation could be monitored over a longer horizon. We only observe two 'treated' years, and it is conceivable that the policy is more successful later on. Our analysis so far does not suggest an increased effect in 2019 compared to 2018. Nevertheless, employees might start seeking out wage information from their employers after hearing success stories of others using this information. If they fear backlash from requesting this information, this fear might diminish after observing that others successfully requested it.

While our experiment focuses on the effect of correcting beliefs about others' wages, future research could take a closer look at whether and how wage transparency can affect wages by spotlighting discriminatory practices. Firms with unequal compensation policies may face public pressure if periodic reporting of gender pay gaps becomes compulsory. Sorting of workers into different firms and industries might then be of particular interest. If wage information is easily accessible, it could reduce gender wage gaps by encouraging firms to increase the wages of women to attract female employees. The current German wage transparency regulation is, given that wage information is hard to access, not able to do so.

APPENDIX

4.A Proofs

A worker and a firm split a pie π. The worker believes he or she can contribute $\hat{c}_i$ to the firm and that the firm pays comparable workers, that is, workers performing comparable tasks, an average wage of $\hat{\hat{w}}_i$. S/he believes that the average contribution of comparable workers to the firm is $\hat{\hat{c}}_i$. The wage in the Nash bargaining solution w_i^* is the w_i characterized by

$$\begin{aligned} \max_{w_i} \quad & \left(w_i - \alpha_i \left(\frac{w_i}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)^2 \right) (\pi - w_i) \\ \text{s.t.} \quad & w_i \geq 0 \\ & \pi \geq w_i \end{aligned}$$

This gives the following objective function

$$L\left(w_i; \alpha_i, \hat{c}_i, \hat{\hat{w}}_i, \hat{\hat{c}}_i, \pi\right) = \left(w_i - \alpha_i \left(\frac{w_i}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)^2 \right) (\pi - w_i) - \lambda(\pi - w_i)$$

The first order conditions for a local maximum are given by

$$\frac{\partial L}{\partial w_i} = -w_i + \alpha_i \left(\frac{w_i}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)^2 + (\pi - w_i) \left(1 - \frac{2\alpha_i \left(\frac{w_i}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)}{\hat{c}_i} \right) + \lambda = 0$$

$$\lambda(\pi - w_i) = 0, \qquad \lambda \geq 0$$

We require $\lambda = 0$, as otherwise we get $L\left(w_i; \alpha_i, \hat{c}_i, \hat{\hat{w}}_i, \hat{\hat{c}}_i, \pi\right) = 0$, which is not a local maximum. Thus, w_i^* is characterized by

$$-w_i^* + \alpha_i \left(\frac{w_i^*}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)^2 + \left(\pi - w_i^*\right) \left(1 - \frac{2\alpha_i \left(\frac{w_i^*}{\hat{c}_i} - \frac{\hat{\hat{w}}_i}{\hat{\hat{c}}_i} \right)}{\hat{c}_i} \right) = 0$$

This gives an implicit function of w_i^* in terms of the agent's beliefs ($\hat{c}_i, \hat{\hat{w}}_i$, and $\hat{\hat{c}}_i$) and aversion to unfair wages (α_i). Solving this expression for w_i^*, we obtain as the only solution that also satisfies the second order condition of $\frac{\partial^2 L}{\partial w_i^2} < 0$:

$$w_i^* = \frac{\pi + \frac{2\hat{c}_i \hat{\hat{w}}_i}{\hat{\hat{c}}_i} + \frac{\hat{c}_i^2}{\alpha_i} - \sqrt{\frac{\hat{c}_i^4}{\alpha_i^2} + \frac{\hat{c}_i^2 (4\hat{c}_i \hat{\hat{w}}_i - \hat{\hat{c}}_i \pi)}{\alpha_i \hat{\hat{c}}_i} + \frac{(\hat{c}_i \hat{\hat{w}}_i - \hat{\hat{c}}_i \pi)^2}{\hat{\hat{c}}_i^2}}}{3} \tag{4.4}$$

Our hypotheses from Section 4.4.1 follow from comparative statics predictions about w_i^* and the assumptions on gender differences in the agent's beliefs ($\hat{c}_i, \hat{\hat{w}}_i$, and $\hat{\hat{c}}_i$) outlined in Section 4.4.1. For the first two results, we assume that information on average wages of comparable others does not affect beliefs $\hat{\hat{c}}$ about average contributions. Information on average performances of comparable others does not affect beliefs about average wages of others. Formally, this means $\frac{\partial \hat{c}_i}{\partial \hat{\hat{w}}_i} = \frac{\partial \hat{\hat{c}}_i}{\partial \hat{\hat{w}}_i} = \frac{\partial \hat{\hat{w}}_i}{\partial \hat{c}_i} = \frac{\partial \hat{\hat{w}}_i}{\partial \hat{\hat{c}}_i} = 0$.

Result 1. *Providing information about a comparable worker's wage increases wages by women compared to men.*

The wage maximizing the Nash product defined in Equation 4.4 has the property that $\frac{\partial w_i^*}{\partial \hat{\hat{w}}_i} > 0$. We assume that women F have pessimistic beliefs about others' wages, so $\hat{\hat{w}}_i^F < \bar{w}$. Men have optimistic beliefs $\hat{\hat{w}}_i^M$, so $\hat{\hat{w}}_i^M > \bar{w}$. After observing information on the correct value $\bar{w}$, beliefs will be updated such that both for men and women $\hat{\hat{w}}_i^F = \hat{\hat{w}}_i^M = \bar{w}$.

Given these assumptions, we consider how making wages transparent (T_w), changes the wage from the Nash bargaining solution for women. We denote this change by $\Delta_{T_w} w_i^{*F}$ and compare this to the change for men, which we denote by $\Delta_{T_w} w_i^{*M}$. This change $\Delta_{T_w} w_i^*$ is defined as the difference in the equilibrium wage if wages are transparent, $w_i^{*T_w}$, compared to when wages are secret, $w_i^{*S_w}$. For this, see that given the assumption $\frac{\partial \hat{c}_i}{\partial \hat{\hat{w}}} = \frac{\partial \hat{\hat{c}}_i}{\partial \hat{\hat{w}}_i} = 0$, we can write

$$\Delta_{T_w} w_i^* = w_i^*\left(\hat{\hat{w}}_i^{T_w}; \cdot\right) - w_i^*\left(\hat{\hat{w}}_i^{S_w}; \cdot\right) = \int_{\hat{\hat{w}}_i^{S_w}}^{\hat{\hat{w}}_i^{T_w}} \underbrace{\frac{\partial w_i^*}{\partial \hat{\hat{w}}_i}}_{>0} d\hat{\hat{w}}_i$$

Here, we use the integral notation to illustrate the dependence of this difference on $\frac{\partial w_i^*}{\partial \hat{\hat{w}}_i}$ and the change in beliefs $\hat{\hat{w}}_i$, which serve as limits of integration.

Since $\bar{w} = \hat{\hat{w}}_i^{T_w} > \hat{\hat{w}}_i^{S_w}$ for women, but $\bar{w} = \hat{\hat{w}}_i^{T_w} < \hat{\hat{w}}_i^{S_m}$ for men, this implies

$$\Delta_{T_w} w_i^{*W} > \Delta_{T_w} w_i^{*M}$$

□

Result 2. *Providing information about a worker's own performance relative to the comparable worker's performance increases wages by women compared to men.*

This proof follows along similar lines as the previous. The wage maximizing the Nash product defined in Equation 4.4 has the property that $\frac{\partial w_i^*}{\partial \hat{c}_i} > 0$ and $\frac{\partial w_i^*}{\partial \hat{\hat{c}}_i} < 0$. Information on c_i and $\bar{c}$ is simultaneously provided. We assume that women have pessimistic beliefs about their performance, denoted by $\hat{c}_i^F$, so $\hat{c}_i^W < c_i$, while men have optimistic beliefs $\hat{c}_i^M > c_i$. After observing information on the correct value c_i, beliefs will be updated such that both for men and women $\hat{c}_i^F = \hat{c}_i^M = c_i$.

Given these assumptions, performance information (T_p) changes the wage from the Nash bargaining solution for women. We denote this change by $\Delta_{T_p} w_i^{*F}$ and compare this to the change for men, which we denote by $\Delta_{T_p} w_i^{*M}$. This change $\Delta_{T_p} w_i^*$ is defined as the difference in the equilibrium wage if performance is transparent, $w_i^{*T_p}$, compared to when performance is secret, $w_i^{*S_p}$.

For this, see that given the assumption $\frac{\partial \hat{\hat{w}}_i}{\partial \hat{c}_i} = \frac{\partial \hat{\hat{w}}_i}{\partial \hat{c}_i} = 0$ we can write

$$\begin{aligned}
\Delta_{T_p} w_i^* &= w_i^*\left(\hat{\hat{c}}_i^{T_p}, \hat{c}_i^{T_p}; \cdot\right) - w_i^*\left(\hat{\hat{c}}_i^{S_p}, \hat{c}_i^{S_p}; \cdot\right) \\
&= \int_{\hat{\hat{c}}_i^{S_p}}^{\hat{\hat{c}}_i^{T_p}} \underbrace{\frac{\partial w_i^*\left(\hat{c}_i = \hat{c}_i^{S_p}\right)}{\partial \hat{\hat{c}}_i}}_{<0} d\hat{\hat{c}}_i + \int_{\hat{c}_i^{S_p}}^{\hat{c}_i^{T_p}} \underbrace{\frac{\partial w_i^*\left(\hat{\hat{c}}_i = \hat{\hat{c}}_i^{T_p}\right)}{\partial \hat{c}_i}}_{>0} d\hat{c}_i
\end{aligned}$$

Ceteris paribus, since $c_i = \hat{c}_i^{T_p} > \hat{c}_i^{S_p}$ for women and $c_i = \hat{c}_i^{T_p} < \hat{c}_i^{S_p}$ for men, this implies

$$\Delta_{T_p} w_i^{*F} > \Delta_{T_p} w_i^{*M}$$

□

Next, we relax the assumption of $\frac{\partial \hat{c}_i}{\partial \hat{\hat{w}}_i} = \frac{\partial \hat{\hat{c}}_i}{\partial \hat{w}_i} = \frac{\partial \hat{\hat{w}}_i}{\partial \hat{c}_i} = \frac{\partial \hat{w}_i}{\partial \hat{\hat{c}}_i} = 0$ and instead posit that $\frac{\partial \hat{\hat{c}}_i}{\partial \hat{\hat{w}}_i} > 0$ and $\frac{\partial \hat{\hat{w}}}{\partial \hat{\hat{c}}_i} > 0$. In this case, if a worker is told that another worker is more productive than anticipated, they will also update beliefs about the wage of the other worker in the same direction.

Result 3. *Providing information about a comparable worker's wage and relative performance jointly has a stronger effect on wages than providing this information separately.*

If $\frac{\partial \hat{\hat{c}}_i}{\partial \hat{\hat{w}}_i} > 0$ and $\frac{\partial \hat{\hat{w}}_i}{\partial \hat{\hat{c}}_i} > 0$, $\Delta_{T_w} w_i^* \neq \int_{\hat{\hat{w}}_i^{S_w}}^{\hat{\hat{w}}_i^{T_w}} \frac{\partial w_i^*}{\partial \hat{\hat{w}}_i} d\hat{\hat{w}}_i$. Instead, we can write that if no performance information is provided, the effect of wage transparency on wages in the

Nash bargaining solution is characterized by

$$\Delta_{T_w} w_i^* = \int_{\hat{\hat{w}}_i^{S_w}}^{\hat{\hat{w}}_i^{T_w}} \frac{\partial w_i^*(\hat{\hat{c}}_i = \hat{\hat{c}}_i^{S_w})}{\partial \hat{\hat{w}}_i} d\hat{\hat{w}}_i + \int_{\hat{\hat{c}}_i^{S_w}}^{\hat{\hat{c}}_i^{T_w}} \underbrace{\frac{\partial w_i^*(\hat{\hat{w}}_i = \hat{\hat{w}}_i^{T_w})}{\partial \hat{\hat{c}}_i}}_{<0} d\hat{\hat{c}}_i$$

$\Delta_{T_w} w_i^*$ is therefore decreasing in $\hat{\hat{c}}_i^{T_w} - \hat{\hat{c}}_i^{S_w}$. Since $\frac{\partial \hat{\hat{c}}_i}{\partial \hat{\hat{w}}_i} > 0$, we know that if and only if $\hat{\hat{w}}_i^{T_w} > \hat{\hat{w}}_i^{S_w}$, also $\hat{\hat{c}}_i^{T_w} > \hat{\hat{c}}_i^{S_w}$ must hold. In other words, providing wage information alone results in a smaller change of equilibrium wages if it also leads to updating of beliefs about performance.

Similarly, the effect of providing performance information is then characterized by

$$\Delta_{T_p} w_i^* = \int_{\hat{\hat{c}}_i^{S_p}}^{\hat{\hat{c}}_i^{T_p}} \frac{\partial w_i^*\left(\hat{c}_i = \hat{c}_i^{S_p}\right)}{\partial \hat{\hat{c}}_i} d\hat{\hat{c}}_i + \int_{\hat{c}_i^{S_p}}^{\hat{c}_i^{T_p}} \frac{\partial w_i^*\left(\hat{\hat{c}}_i = \hat{\hat{c}}_i^{T_p}\right)}{\partial \hat{c}_i} d\hat{c}_i + \int_{\hat{\hat{w}}_i^{S_p}}^{\hat{\hat{w}}_i^{T_p}} \underbrace{\frac{\partial w_i^*(\hat{\hat{c}}_i = \hat{\hat{c}}_i^{T_p})}{\partial \hat{\hat{w}}_i}}_{>0} d\hat{\hat{w}}_i$$

$\Delta_{T_p} w_i^*$ is therefore increasing in $\hat{\hat{w}}_i^{T_p} - \hat{\hat{w}}_i^{S_p}$. Since $\frac{\partial \hat{\hat{w}}_i}{\partial \hat{\hat{c}}_i} > 0$, we know that if and only if $\hat{\hat{c}}_i^{T_p} > \hat{\hat{c}}_i^{S_p}$, also $\hat{\hat{w}}_i^{T_p} > \hat{\hat{w}}_i^{S_p}$ must hold.

Note that if wage and performance information are provided jointly, we are back in the cases considered in Result 1 and Result 2, as the respective beliefs about wages or performance will be fixed.

This implies that the joint effect of providing wage and performance information on equilibrium wages is larger for women than the sum of the effects of providing the two types of information separately. Given $\hat{\hat{w}}_i^{T_w} > \hat{\hat{w}}_i^{S_w}$ and $\hat{\hat{c}}_i^{T_p} < \hat{\hat{c}}_i^{S_p}$, $\Delta_{T_w} w_i^*$ and $\Delta_{T_p} w_i^*$ are smaller if provided separately than if provided jointly. For men, the reverse holds true. The effect is thus muted if the information is provided separately compared to provided simultaneously.

□

4.B Prolific pre-study

We conducted a pre-study before running the experiment described in Section 4.4.2. This pre-study is designed to inform us on which tasks are perceived to favor male participants. We recruited 100 participants on Prolific. We selected participants from the Netherlands in an age bracket from 18 to 30 years to match the subject pool from the University of Amsterdam.

The survey asks participants whether the average number of correctly solved tasks was 5% higher for men, 5% higher for women or the average numbers of correctly solved tasks of men and women were within 5% of each other.

We asked participants about their estimates about men's and women's performance in three tasks. The first two tasks are the maze and the matrix task, described in Section 4.4.2. The third task are Raven's matrices.

On top of a one Pound base payment, we use a bonus payment to incentivize this task. If the participant's answer matches the results of a corresponding experimental study, the participant receives 30 pence per correct answer. To incentivize accurate beliefs in the matrix task, we use Schram et al. (2019) for the matrix task, Gneezy et al. (2003) for the maze task, and Crucian and Berenbaum (1998) for the Raven's matrices.

Table 4.B.1 provides the shares of respondents who believe that men or who believe that women solve at least 5% more tasks correctly.

	Raven's task	Matrix task	Maze task
Men	27%	45%	33%
Women	28%	20%	23%

Table 4.B.1: Overview of pre-study results

For Raven's matrices, we see that there is an almost equal share of participants that believe that men versus women perform better in this task (27% versus 28%, respectively). These shares are not significantly different (t-test; $p = 0.894$).

45% of the respondents believe that men solve significantly more elements correctly in the matrix task, while only 20% believe that women do so. This difference is significant (t-test; $p = 0.002$).

The pattern is similar for the maze task. Here, 33% of the respondents believe that men perform better, 23% believe that women perform better. While these shares differ by 43%, this difference is not statistically significant (t-test; $p = 0.183$).

Given this evidence, we do not include Raven's matrices in our experiment, as this task does not appear to respondents as favoring male participants.

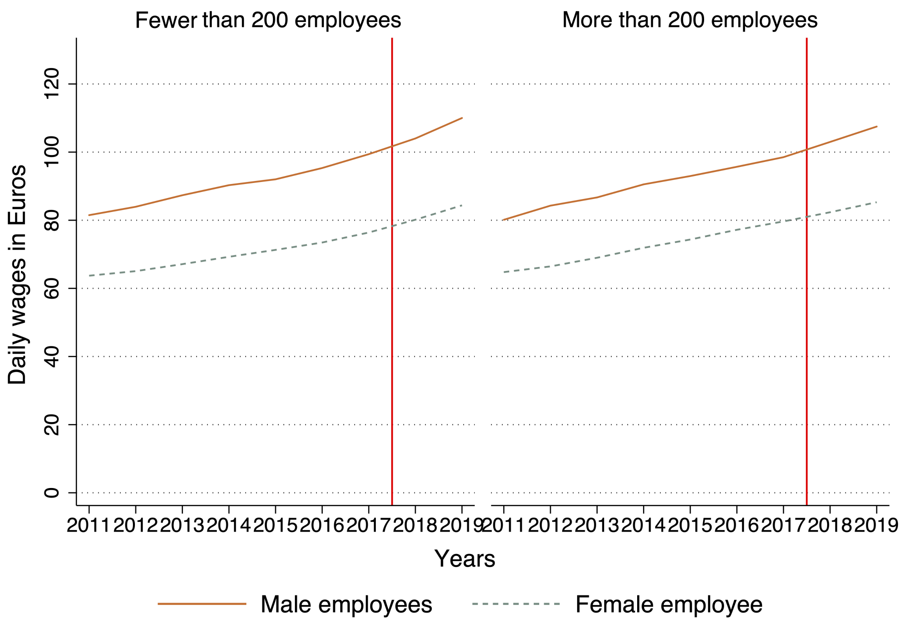
4.C Additional analyses of field data

This appendix complements the analysis from Section 4.3. We will first give additional tables and figures using LIAB data, then provide the main analysis using data from SIEED, and finally present some heterogeneity analysis.

4.C.1 Additional results using LIAB

This Figure 4.C.1 presents the unadjusted daily wages by gender in firms with fewer than 100 employees (left) or more than 200 employees (right), considering firms with 150 to 250 employees. Table 4.C.1 provides Diff-in-Diff estimates of the effect of the wage transparency regulation on the employee's propensity to change firms, see Section 4.3.4. Figure 4.C.2 shows the share of firms by firm size in our sample, around the cutoff of 200 employees.

Tables 4.C.2, 4.C.5, 4.C.6, 4.C.7, and 4.C.8 provide the regression analyses for the robustness checks outlined in Section 4.3.5. The main results of these regressions are also illustrated in Figure 4.1.



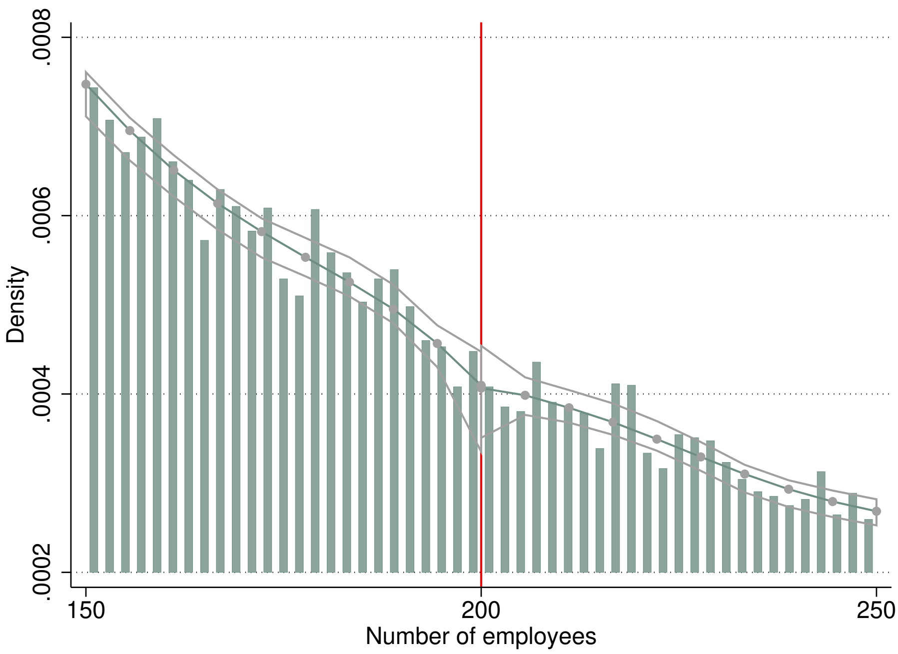
Notes: Raw data of daily wages from 2011 to 2019 by gender and by firm size. Includes observations from firms with 150 to 250 employees in 2018. The red vertical line indicates the introduction of the wage transparency regulation.

Figure 4.C.1: The gender gap in wages in firms with fewer vs. in firms with at least 200 employees

	Indicator of employment change					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Large × Post	-0.0003	0.0001	0.0046	0.0034	0.0041	0.0063
	(-0.02)	(0.01)	(0.34)	(0.28)	(0.33)	(0.55)
Female × Large × Post	0.0050			0.0033		
	(0.43)			(0.30)		
Female × Large	0.0022			0.0391^{**}		
	(0.07)			(2.27)		
Female × Post	-0.0093			-0.0192^{***}		
	(-1.31)			(-2.88)		
Ind. time-varying controls	✓	✓	✓			
Firm FE	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	493,325	274,877	217,797	663,787	371,407	291,543

Notes: Impact of transparency regulation on an indicator variable equal to one if the employee changes the establishment they work at by the next year and zero otherwise. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.1: Diff-in-Diff estimates of impact of wage transparency law on the propensity of employees to seek alternative employment



Notes: Plot of the share of firms by firm size, measured by the number of employees, in the range of 150 to 250, split by the cutoff of 200 (red vertical line). The center line indicates the estimated density, the gray lines indicate the 95% confidence interval around this.

Figure 4.C.2: Density plot of the firm size

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Diff-in-disc	0.0307	0.0320	0.0008	0.0024	0.0025	-0.0578
	(0.88)	(0.96)	(0.02)	(0.04)	(0.04)	(-1.11)
Female × Diff-in-disc	-0.0251			-0.0603		
	(-0.70)			(-1.02)		
Ind. time-varying controls	✓	✓	✓			
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	639,395	357,630	281,765	852,465	478,000	374,465

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-discontinuity specification. Time-varying controls include age squared, education and an indicator for part-time workers. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.2: Diff-in-Disc estimates of impact of wage transparency law on daily wages

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Large × Post	0.0055	0.0050	-0.0017	0.0040	0.0112*	0.0011
	(1.18)	(1.07)	(-0.26)	(0.64)	(1.79)	(0.16)
Female × Large × Post	-0.0073			-0.0064		
	(-1.12)			(-0.82)		
Female × Large	-0.1361***			-0.0481**		
	(-3.88)			(-2.10)		
Female × Post	0.0187***			0.0060		
	(4.24)			(1.25)		
Ind. time-varying controls	✓	✓	✓			
Firm FE	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	585,822	333,183	252,051	778,441	446,733	340,632

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-difference specification, using firm sizes recorded in 2017. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.3: Diff-in-Diff estimates for impact of wage transparency on daily wages, based on firm size in 2017

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Diff-in-disc	0.0168	0.0150	0.0214	0.0014	0.0012	-0.0029
	(0.60)	(0.54)	(0.88)	(0.04)	(0.03)	(-0.09)
Female × Diff-in-disc	0.0025			-0.0044		
	(0.09)			(-0.12)		
Ind. time-varying controls	✓	✓	✓			
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	642,205	365,685	276,520	863,855	490,397	373,458

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-discontinuity specification, using firm sizes recorded in 2017. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.4: Diff-in-Disc estimates for impact of wage transparency on daily wages, based on firm size in 2017

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Large × Post	0.0023	0.0009	0.0028	0.0050	0.0043	0.0022
	(0.47)	(0.19)	(0.42)	(0.79)	(0.69)	(0.32)
Female × Large × Post	-0.0001			-0.0026		
	(-0.01)			(-0.34)		
Female × Large	-0.0231			0.0045		
	(-0.76)			(0.20)		
Female × Post	0.0145***			0.0041		
	(3.25)			(0.80)		
Ind. time-varying controls	✓	✓	✓			
Firm FE	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	576,495	319,700	256,186	770,238	428,867	340,589

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Sample excludes employment spells with top-coded observations. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.5: Diff-in-Diff estimates of impact of wage transparency law on daily wages excluding top-coded observations

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Diff-in-disc	0.0326	0.0333	0.0029	0.0074	0.0074	-0.0543
	(0.93)	(0.99)	(0.08)	(0.13)	(0.13)	(-1.07)
Female × Diff-in-disc	-0.0252			-0.0617		
	(-0.70)			(-1.05)		
Ind. time-varying controls	✓	✓	✓			
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	632,080	351,645	280,435	844,495	471,468	373,027

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Sample excludes employment spells with top-coded observations. Estimates from difference-in-discontinuity specification. Time-varying controls include age squared, education and an indicator for part-time workers. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.6: Diff-in-Disc estimates of impact of wage transparency law on daily wages excluding top-coded observations

	Log of daily wage				
	(1)	(2)	(3)	(4)	(5)
Large × Post	0.0009	-0.0008	0.0022	0.0031	0.0015
	(0.22)	(-0.19)	(0.46)	(0.56)	(0.22)
Female × Large × Post	0.0022	0.0023	-0.0001	-0.0003	0.0013
	(0.41)	(0.39)	(-0.01)	(-0.04)	(0.16)
Female × Large	-0.0425*	-0.0442	-0.0249	-0.0283	-0.0208
	(-1.87)	(-1.42)	(-0.83)	(-0.71)	(-0.38)
Female × Post	0.0152***	0.0164***	0.0146***	0.0123**	0.0144**
	(4.38)	(4.21)	(3.30)	(2.45)	(2.28)
Ind. time-varying controls	✓	✓	✓	✓	✓
Firm FE	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Firm size	130-270	140-260	150-250	160-240	170-230
Observations	852,267	707,938	584,026	464,504	333,935

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.7: Diff-in-Diff with different bandwidths

	Log of daily wage				
	(1)	(2)	(3)	(4)	(5)
Diff-in-disc	0.0289	0.0279	0.0307	0.0348	0.0573
	(1.07)	(0.91)	(0.88)	(0.85)	(1.24)
Female × Diff-in-disc	-0.0246	-0.0117	-0.0251	-0.0392	-0.0586
	(-0.86)	(-0.36)	(-0.70)	(-0.97)	(-1.29)
Time FE	✓	✓	✓	✓	✓
Firm size	130-270	140-260	150-250	160-240	170-230
Observations	926,022	772,753	639,395	508,662	368,058

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-discontinuity specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.8: Diff-in-Disc with different bandwidths

4.C.2 Results from SIEED

Our secondary data source is the German Sample of Integrated Employer-Employee Data (SIEED). This employer-employee matched administrative data set covers 1.5% of all German establishments and contains information on employment spells of all employees. Employee-level demographic information includes age, completed education and whether the work was part-time. Data at the establishment level, including the total number of employees, are obtained from the linked Establishment-History-Panel (BHP). A detailed description of SIEED can be found in Schmidtlein et al. (2020).

We observe employment spells from 2011 to 2018. We again discard all observations with a zero wage, indicating employment interruptions. This leaves 1,842,584 relevant observations from 544,437 individuals at 16,049 firms in our main sample, substantially more than in our primary analysis. Table 4.C.9 reports summary statistics for this data set, Table 4.C.10 the Diff-in-Diff analysis, Table 4.C.11 the Diff-in-Disc analysis and Figure 4.C.3 provides the event study specification.[33]

[33] Source DOI: 10.5164/IAB.FDZD.2014.en.v1, own calculations. We use these data for all results in Section 4.C.2.

	Men		Women	
	Large firms (1)	Small firms (2)	Large firms (3)	Small firms (4)
Daily Wage	98.18	96.96	67.75	67.65
	(54.18)	(52.62)	(45.88)	(45.25)
Age	43.66	43.56	43.67	44.00
	(11.98)	(12.01)	(11.80)	(11.80)
College educated	17.34%	16.18%	15.87%	15.72%
Part-time	16.00%	15.20%	53.73%	52.99%
Firms	5,755	10,162	5,623	9,840
Individuals	126,111	179,476	106,102	152,222
Observations	415,813	594,111	340,610	492,050

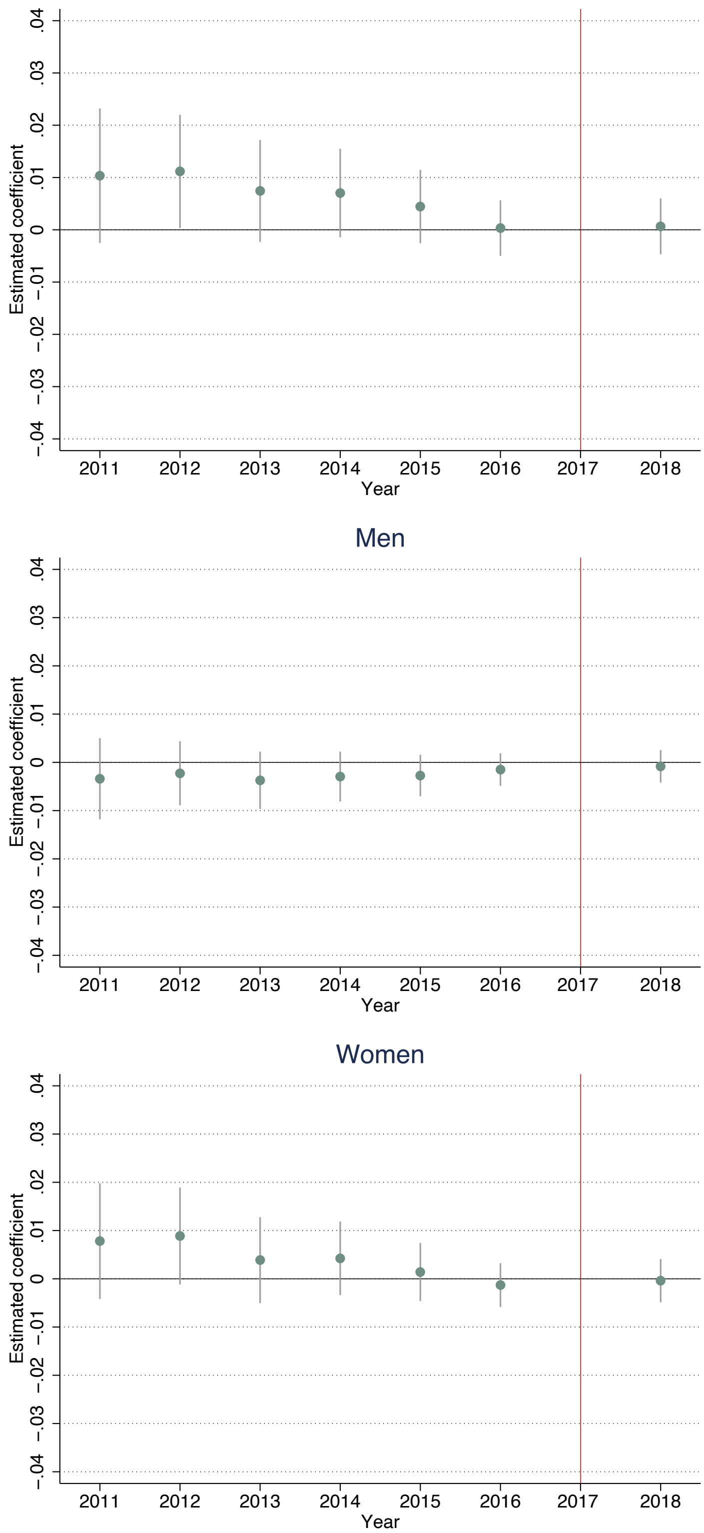
Notes: This table reports unconditional means and standard deviations in parentheses of key variables for individuals in large and small firms, split by gender. The descriptive statistics include all data in our SIEED panel from 2011 to 2018 in firms with 150 to 250 employees in 2018. 'Age' refers to the employee's age in years, 'College educated' is an indicator of whether the employee has at least some university or university of applied sciences education, and 'Part-time' is an indicator of whether the employee works part time.

Table 4.C.9: Summary statistics using SIEED

	Log of daily wage					
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)
Large × Post	0.0006	0.0009	-0.0024	0.0023	0.0023	0.0018
	(0.0022)	(0.0022)	(0.0030)	(0.0027)	(0.0026)	(0.0030)
Female × Large × Post	-0.0032			-0.0004		
	(0.0034)			(0.0036)		
Female × Large	-0.0134			0.021		
	(0.0158)			(0.0142)		
Female × Post	0.0213***			0.0139***		
	(0.0021)			(0.0022)		
Ind. time-varying controls	✓	✓	✓			
Firm FE	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250
Observations	1,137,638	632,974	504,269	1,652,424	909,136	742,997

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2018 in SIEED. Standard errors are clustered at the firm level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.10: Diff-in-Diff estimates using SIEED



Notes: Event study analysis of the impact of wage transparency regulation on log of daily wage. The top figure provides the differential impact for women vs. men, the bottom two figures separate event study specifications. Firms with more than 200 employees are classified as treated. Individual-, firm- and year-fixed effects are included. Time varying controls include age squared, education and part-time workers. 1,137,638 observations, including both men and women. Error bars indicate the 95% confidence interval. Standard errors are clustered at the firm level.

Figure 4.C.3: Gender-specific effects of the transparency law

	Log of daily wage		
	Both gender (1)	Men (2)	Women (3)
Diff-in-disc	-0.0028 (0.0209)	-0.0028 (0.0209)	-0.0024 (0.0221)
Female × Diff-in-disc	0.0002 (0.0192)		
Time FE	✓	✓	✓
Firm size	150-250	150-250	150-250
Observations	1,833,178	1,006,963	826,215

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually. Estimates from difference-in-discontinuity specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2011 to 2018 in SIEED. Standard errors are clustered at the firm level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.11: Diff-in-Disc estimates using SIEED

4.C.3 Heterogeneity analysis

This section presents the analysis by collective bargaining status, as outlined in Section 4.3.4. Table 4.C.12 provides results for firms covered by collective bargaining agreement, Table 4.C.13 for firms not covered by such agreements.

	Log of daily wage						
	Both gender (1)	Men (2)	Women (3)	Both gender (4)	Men (5)	Women (6)	Both gender (7)
Large × Post	0.0025	0.0009	0.0064	0.0060	0.0058	0.0137	0.0280
	(0.25)	(0.09)	(0.65)	(0.45)	(0.45)	(1.22)	(1.01)
Female × Large × Post	0.0026			0.0078			-0.0431
	(0.22)			(0.53)			(-1.48)
Female × Large	0.0140			0.0001			
	(0.06)			(0.00)			
Female × Post	0.0097			0.0000			0.0219
	(1.19)			(0.01)			(0.87)
Ind. time-varying controls	✓	✓	✓				✓
Firm FE	✓	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250	150-250
Observations	212,959	113,069	99,889	260,034	136,022	124,011	45,479

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually in establishments bound by an industry-wide or firm-level wage agreement in columns (1) - (6). Column (7) also considers establishments that are not bound by collective bargaining agreements, but base their wages on these agreements. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2010 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.12: Diff-in-Diff estimates in establishments bound by a wage agreement

	Log of daily wage						
	Both gender	Men	Women	Both gender	Men	Women	Both gender
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Large × Post	0.0035	0.0018	-0.0236	0.0131	0.0132	-0.0305	-0.0295*
	(0.14)	(0.08)	(-0.95)	(0.38)	(0.38)	(-1.30)	(-1.75)
Female × Large × Post	-0.0219			-0.0436			-0.0008
	(-0.90)			(-1.37)			(-0.03)
Female × Large	0.0959*			0.1308			0.1136*
	(1.79)			(1.09)			(1.95)
Female × Post	0.0235			0.0078			0.0278**
	(1.32)			(0.43)			(2.05)
Ind. time-varying controls	✓	✓	✓				✓
Firm FE	✓	✓	✓	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓	✓	✓
Firm size	150-250	150-250	150-250	150-250	150-250	150-250	150-250
Observations	68,156	39,446	28,710	88,631	53,308	35,322	22,662

Notes: Impact of transparency regulation on the gender wage gap and the wages of men and women individually in establishments not bound by an industry-wide or firm-level wage agreement in columns (1) - (6). Column (7) only considers establishments that also do not base their wages on these agreements. Estimates from difference-in-difference specification. Individual time-varying controls include age squared, education and part-time occupation. Includes observations from 2010 to 2019. Standard errors are clustered at the firm level. T-statistics in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.C.13: Diff-in-Diff estimates in establishments not bound by a wage agreement

4.D Additional analyses of laboratory data

	Worker's wage	
	(1)	(2)
Worker contribution	0.55***	0.48***
	(0.04)	(0.03)
Firm contribution	0.24**	0.24**
	(0.09)	(0.10)
Endo wage	-10.81	
	(18.82)	
Exo wage	21.52	
	(18.95)	
Endo wage × Worker contribution	0.02	
	(0.05)	
Exo wage × Worker contribution	-0.03	
	(0.06)	
Performance		-41.81**
		(16.98)
Performance × Worker contribution		0.14***
		(0.05)
Constant	-1.71	20.47
	(35.84)	(35.42)
Part FE	✓	✓
Period FE	✓	✓
Laboratory FE	✓	✓
Observations	1548	1548
Clusters	66	66
R-squared	0.265	0.268

Notes: Results are from ordinary least squares regression of the worker's wage, restricting the sample to periods in which subjects enter negotiations. Worker contribution is a control for the worker's contribution to the negotiation pie, Firm contribution for the firm's contribution to the negotiation pie. Endo wage and Exo wage are indicators of whether wage information was provided endogenously or exogenously, respectively. Female indicates whether a participant is female. Performance is an indicator of whether information of the workers' performances is provided. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.D.1: The interaction of wage and performance information with the worker's contribution

	Worker's decision to opt out of negotiations				
	(1)	(2)	(3)	(4)	(5)
Worker contribution	-0.034***	-0.037***	-0.035***	-0.036***	-0.034***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.007)
Female	0.026*		0.004		0.028**
	(0.013)		(0.015)		(0.014)
Wage info		0.033***	0.016		
		(0.012)	(0.013)		
Wage info × Female			0.033		
			(0.025)		
Performance info				0.009	0.011
				(0.010)	(0.010)
Performance info × Female					-0.004
					(0.016)
Constant	0.125***	0.126***	0.116***	0.141***	0.119***
	(0.027)	(0.027)	(0.028)	(0.029)	(0.025)
Part FE	✓	✓	✓	✓	✓
Period FE	✓	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓	✓
Observations	1546	1546	1546	1546	1546
Clusters	66	66	66	66	66
R-squared	0.061	0.063	0.070	0.058	0.062

Notes: Results are from OLS regression of the participant's (binary) decision to opt out of negotiations. Worker contribution is a control for the worker's contribution to the negotiation pie (in hundred units). Female indicates whether a participant is female. Wage info is an indicator of whether wage information was (potentially) provided. Performance is an indicator of whether information of the workers' performances is provided. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.D.2: The effect of information on opting out of negotiations

	Worker's wage							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Worker contribution	0.53***	0.53***	0.53***	0.53***	0.53***	0.54***	0.53***	0.53***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Firm contribution	0.25**	0.25**	0.24**	0.25**	0.24**	0.24**	0.25**	0.25**
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)
Endo wage	-0.04			-2.46			1.11	-2.84
	(10.06)			(14.00)			(12.86)	(17.79)
Exo wage	14.37	14.39*		15.19			18.63	24.01
	(8.99)	(7.93)		(12.65)			(11.63)	(16.06)
Female			5.79	4.88		8.65		9.90
			(6.04)	(12.21)		(9.09)		(18.02)
Endo wage × Female				4.90				7.85
				(14.96)				(22.68)
Exo wage × Female				-1.67				-11.43
				(15.75)				(22.49)
Performance					11.82**	14.63*	15.36*	20.59
					(5.31)	(7.49)	(8.48)	(13.24)
Performance × Female						-5.78		-10.77
						(9.94)		(20.46)
Performance × Endo wage							-2.17	0.36
							(12.87)	(18.01)
Performance × Exo wage							-8.49	-18.08
							(12.90)	(18.73)
Performance × Endo wage × Female								-5.24
								(25.03)
Performance × Exo wage × Female								20.22
								(25.57)
Constant	3.17	3.15	5.86	-1.23	2.36	-4.00	-5.81	-12.44
	(36.91)	(36.09)	(37.87)	(39.25)	(37.75)	(38.90)	(36.57)	(39.82)
Part FE	✓	✓	✓	✓	✓	✓	✓	✓
Period FE	✓	✓	✓	✓	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1486	1486	1486	1486	1486	1486	1486	1486
Clusters	66	66	66	66	66	66	66	66
R-squared	0.247	0.247	0.245	0.248	0.246	0.247	0.249	0.251

Notes: Results are from ordinary least squares regression of the worker's wage, restricting the sample to periods in which subjects enter negotiations. Worker contribution is a control for the worker's contribution to the negotiation pie, Firm contribution for the firm's contribution to the negotiation pie. Endo wage and Exo wage are indicators of whether wage information was provided endogenously or exogenously, respectively. Female indicates whether a participant is female. Performance is an indicator of whether information of the workers' performances is provided. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.D.3: The effect of wage and performance information conditional on negotiation entry

	Worker's wage				
	(1)	(2)	(3)	(4)	(5)
Worker contribution	0.57***	0.70***	0.61***	0.50***	0.55***
	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)
Firm contribution	0.06	0.07	0.23*	0.25**	0.35***
	(0.18)	(0.18)	(0.13)	(0.12)	(0.10)
Info choice	-17.77	76.86**			
	(11.72)	(30.23)			
Info choice × Worker contribution		-0.25***			
		(0.08)			
Endo wage			4.79		
			(11.80)		
Exo wage				26.14**	23.32
				(11.69)	(19.38)
Exo Wage × Worker contribution					-0.03
					(0.06)
Constant	67.34	19.77	-17.90	4.63	-41.58
	(66.20)	(69.78)	(48.85)	(42.52)	(39.47)
Part FE	✓	✓	✓	✓	✓
Period FE	✓	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓	✓
Sample	*EndoWage*	*EndoWage*	*No wage info*	*Wage info*	*NoWage & ExoWage*
Observations	515	515	789	759	1033
Clusters	22	22	44	44	44
R-squared	0.272	0.284	0.303	0.240	0.272

Notes: Results are from ordinary least squares regression of the worker's wage. Worker contribution is a control for the worker's contribution to the negotiation pie, Firm contribution for the firm's contribution to the negotiation pie. Endo wage and Exo wage are indicators of whether wage information was provided endogenously or exogenously, respectively. Info choice indicates whether the participant chose to receive wage information. Standard errors are clustered at the matching-group level and shown in parentheses. Sample refers to the treatment(s) from which the observations for the analysis stem; *No Wage info* refers to observations from *NoWage* and individuals choosing no wage information in *EndoWage*, *Wage info* refers to observations from *ExoWage* and individuals choosing wage information in *EndoWage*.

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

Table 4.D.4: Effects of requesting wage information on wages

	Difference in beliefs			
	(1)	(2)	(3)	(4)
Error in belief of other's wage		-0.02**		
		(0.01)		
Error in belief of other's performance				-7.25***
				(2.56)
Constant	2.89	4.36	43.96***	44.49***
	(7.63)	(8.00)	(15.07)	(14.97)
Part FE	✓	✓	✓	✓
Laboratory FE	✓	✓	✓	✓
Observations	144	144	128	126
Clusters	44	44	41	41
R-squared	0.012	0.020	0.039	0.077

Notes: Results are from ordinary least squares regression of the difference in beliefs between *Elicitation 2* and *Elicitation 3*. Error in belief of other's wage is defined as the difference between the subject's beliefs about the comparable worker's wage and the comparable worker's actual wage. Error in belief of other's performance is defined as the difference between the subject's beliefs about the comparable worker's number of correctly solved in the production tasks and the comparable worker's actual number correctly solved elements. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.D.5: Changes in beliefs between *Elicitation 2* and *Elicitation 3*

	Negotiation breakdown	
	(1)	(2)
Worker contribution	-0.01	-0.00
	(0.01)	(0.01)
Firm contribution	-0.04*	-0.05**
	(0.02)	(0.02)
Optimistic	0.06***	
	(0.02)	
Overconfident		-0.00
		(0.02)
Constant	0.27***	0.30***
	(0.09)	(0.09)
Part FE	✓	✓
Period FE	✓	✓
Laboratory FE	✓	✓
Observations	1548	1545
Clusters	66	66
R-squared	0.026	0.014

Notes: Results are from ordinary least squares regression of an indicator that negotiations broke down and resulted in zero payoff for worker and firm. Worker contribution is a control for the worker's contribution to the negotiation pie (in hundred units), Firm contribution for the firm's contribution to the negotiation pie (in hundred units). Optimist indicates that a subject's beliefs about the comparable worker's wage are too optimistic, Overconfident indicates that a subject's beliefs about his or her own performance relative to the comparable worker's are too optimistic. Standard errors are clustered at the matching-group level and shown in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.D.6: Negotiation breakdown by type

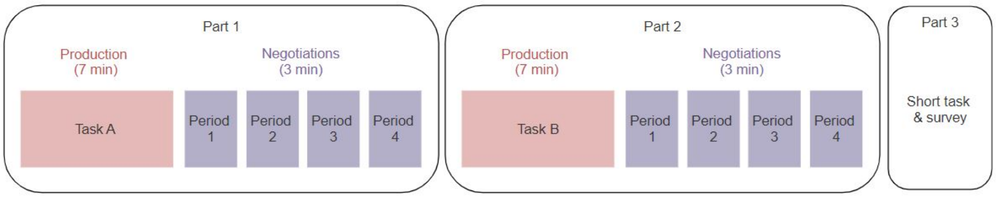
4.E Experimental instructions

Instructions

Please read these instructions carefully.

If you follow the instructions carefully, you may earn a considerable amount of money. Your earnings will depend on your decisions and may depend on other participants' decisions as well as chance.

This experiment consists of 3 parts. In the first two parts, a firm and a worker negotiate a wage for the worker for producing output. First, the worker produces an output. Then, firms and workers will be randomly matched and negotiate over a wage for the worker. There will be **4 negotiation rounds in every part** after each production stage. In part 3, there will be a short task and a survey. Part 3 is independent of part 1 and 2. The graph below shows the flow of the experiment.



Production stage

The production stage determines the total number of points that workers and firms can split during negotiations, called the budget. The budget is determined by the sum of the firm's and the worker's contributions. The firm knows the size of the budget, the worker does not. It is generated as follows:

Firm For each part, every firm draws a random number between between 3000 points and 450 points as the **fixed firm contribution**. This fixed contribution cannot be influenced by the firm and remains the same for the firm during a part. Each firm has a different draw for the firm contribution.

Worker Every worker has to perform a task. They are asked to solve as many elements as possible in seven minutes. At a later stage, more detailed instructions about these tasks will be provided. There will be different tasks for part 1 and part 2. We will call the number of elements solved in a task the **worker's performance**. The more elements the worker solves, so the higher the worker's performance is, the more points can be split between the worker and the firm. The worker increases the budget by by **35/20 points in part 1** for each correctly solved element, and by **35/20 points in part 2**.

Although firms do not have to perform the tasks, they will be shown the task that workers have to perform. The performance of firms in these tasks does not have any consequences for the budget or anyone's payoff.

In sum, the budget is the number of correctly solved elements by the worker multiplied by 35/20 (part 1) or 35/20 (part 2), plus the fixed contribution by the firm.

Negotiation stage

There are four negotiation periods in each negotiation stage. Every period, a worker and a firm are randomly paired and negotiate to split the budget they generated together. Within a part, you have a new negotiation partner in each period. This means that if a firm and a worker are paired in a period, they will be paired with someone else the next period of that part. Since a new pair is formed each period, the budget that can be split between a firm and a worker differs from period to period.

Timeline In period 1, all workers automatically enter negotiations. At the start of the subsequent periods 2-4, the worker must decide whether or not to enter negotiations. If the worker decides not to enter negotiations, the worker will receive 150 points and the firm receives the remainder. If the worker does not enter negotiations, both the worker and the firm will have to wait for other pairs, while they negotiate; the new period starts after all negotiations have ended.

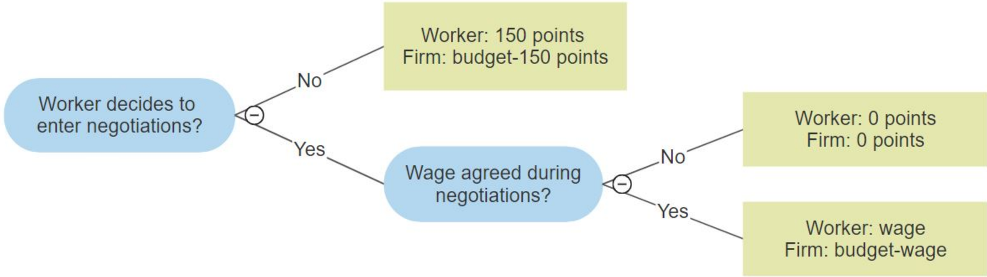
In period 1 and in later periods, if the worker decides to enter negotiations:

1. Both the worker and the firm submit an **initial wage proposal**. The firm can offer to the worker a wage between zero and the budget generated for this period. The worker can request a positive wage. These initial wage proposals are not binding. Workers and firms still need to submit binding wage proposals later on.

2. The worker and the firm have 3 minutes to negotiate a wage. For the negotiation, they will use a **chat**. No personally identifiable information such as names, age or gender is allowed in the chat. They can enter binding wage proposals in a separate entry field. If they agree on a wage proposed by their negotiation partner, they can **click on 'accept'**.

3. If a wage agreement is reached, this wage is implemented. If there is no agreement, that is, neither the firm nor the worker accepted the other's wage proposal, both the worker and the firm receive a payoff of zero points for this period.

See below for a graphical outline of the negotiation stage.



What do you know when you negotiate?

Workers and firms have different information when negotiating. Workers also have different information in part 1 and part 2.

After the first period of each part, the worker has certain information about a **comparable worker**. In periods 2,3 and 4, the comparable worker is always the worker that was paired to the same firm in period 1 as the worker is paired to in the current period. The comparable worker did the same production task.

For example, if worker A was paired to firm X in period 1 and worker B gets paired to firm X in period 2, then worker A will be the comparable worker for worker B in period 2. If worker C is paired to firm X in period 3, worker A will be the comparable worker for worker C in period 3.

Performance information [ORDER DEPENDS ON TREATMENT:

In part 1, the worker and the firm know the worker's performance in this part's production task as well the comparable worker's performance in the same production task.

In part 2, neither the worker nor the firm receive any information about the worker's performance in that part's production task. The worker and the firm also do not know the comparable worker's performance.]

Wage information [ENDOWAGE: In both part 1 and part 2, the worker can decide whether he or she wants to receive information on the comparable worker's wage. Buying this information costs 10 points. **If the worker acquires information, he or she will be told the wage that the comparable worker received**. This is the wage that the firm with which the worker is currently paired to paid another worker in the first period.]

[NOWAGE: In both part 1 and part 2, **the worker does not know the wage of the comparable worker**.]

[EXOWAGE: In both part 1 and part 2, **the worker will be told the wage that the comparable worker received**. This is the wage that the firm with which the worker is currently paired to paid another worker in the first period.]

There is no information on the firm's fixed contribution. As stated before, workers do not know the size of the budget that can be split in each period.

In contrast, the firm always knows the size of the budget. Firms also know all other information that is provided to the worker, including information about the comparable worker.

Payoff summary

For this experiment, you will be paid a show-up fee of 6 Euros. Additionally, you will be paid based on your decisions in the experiment.

To summarize, the payoffs for the worker and the firm in a period are the following:

- If the worker does not enter negotiations: 150 points for the worker, the budget minus 150 points for the firm.

- If the worker enters negotiations: The agreed upon wage for the worker and the budget minus the wage for the firm if an agreement is reached, zero points for both worker and firm if no agreement is reached.

One period from either part 1 or part 2 is randomly selected for payment. All periods are equally likely to be selected. Your decisions do not have any influence on the probability that a certain period is selected for payment.

[WORKER: Furthermore, you will be paid based on your estimate of wages and of performances and for the short task in part 3. You will receive detailed information about the payment of these task later on. You will also receive 4 Euros for completing the questionnaire at the end.]

At the end of the experiment, points will be converted to Euros. 25 points will be converted to one Euro. So each point is worth 0.04 Euros.

Your role

You will have the role of [WORKER: **a worker**] [FIRM: **a firm**].

Instruction Summation Task

In this task you have to find the largest numbers in two different matrices and sum them up.

Each element contains two matrices. Every matrix contains exactly 49 numbers, displayed in seven rows and seven columns. The numbers are randomly generated by the computer. First, find the largest number in each of the two matrices. Then, find the sum of these two numbers and enter your answer.

As an example, see the two matrices below. In the left matrix, the largest number is 85. In the right matrix, the largest number is 79. The sum of 85 and 79 is 164. The correct answer for this example is therefore 164.

17	59	23	31	11	35	53
40	53	57	11	18	61	20
42	84	12	29	43	45	28
29	23	33	45	30	25	38
20	24	85	15	72	21	47
36	36	16	58	45	16	26
76	15	60	52	29	14	26

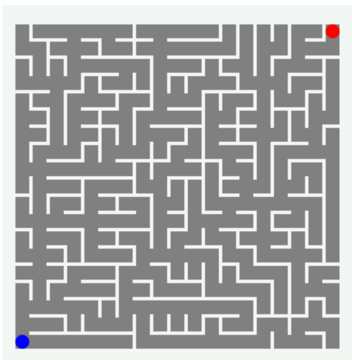
17	39	22	21	34	16	41
74	79	31	13	31	21	13
22	19	17	16	27	41	14
51	58	60	17	40	60	17
27	37	50	79	75	19	34
78	47	51	75	25	58	16
12	10	21	15	70	44	28

Your goal is to solve as many elements as you can within 7 minutes (you can answer up to 50 questions in total). For every question that you solve correctly, the negotiation-stage budget is increased by 35 points.

Instruction Maze Task

In this task you must navigate through a maze. Your current position in indicated by a blue dot, which always starts in the bottom-left corner of the maze. The end of the maze is indicated by a red dot, which always appears in the upper-right corner of the maze.

You can move the blue dot using the arrow keys on your keyboard. Walls of the maze are shown in white. An example maze is shown below.



Your goal is to solve as many mazes as you can within 7 minutes by moving the blue dot onto the exit marked in red. For every maze that you solve, the negotiation-stage budget is increased by 20 points.

Belief elicitation[34]

Estimates about performance

Please provide an estimate of your performance and the performance of another (randomly chosen) worker in the [summation task] [maze task]. Please enter below how many elements you think that you and the randomly chosen worker solved correctly.

At the end of the experiment, one of the questions about your estimates will be chosen for payment. You will receive a bonus of 3 Euros if your guess is close enough to the actual answer. It is in your interest to provide accurate guesses, as this increases the probability of receiving the bonus. If you would like to know more about the mechanism we use to determine whether you receive this bonus, feel free to click on the button below.

Optional: Click here for information about the mechanism

[IF CLICKED: If a question is chosen for payment, the probability that you receive a bonus payment of 3 Euros will depend on your prediction error. This prediction error is the distance between your estimate and the correct number. The closer your estimate is to the correct answer, the larger is the probability that you will receive the bonus.

Assume that your actual performance is X solved [summations] [mazes] and you guessed that you had Y solved [summations] [mazes]. In this case your squared prediction error is $(X-Y)^2$. To determine the probability of receiving the bonus, the computer first draws a number between 0 and 20, let's call this number T. Then this number T is compared to your squared prediction error. If T is larger than the squared error, you will receive the bonus payment for this question. If your squared prediction error is larger than or equal to T, you will not receive a bonus for this question.]

Estimates of wage of others

Please provide an estimate of the wage of another (randomly chosen) worker in the [summation task] [maze task]. Please enter below how many points you think that the randomly chosen worker received in the last negotiation period.

At the end of the experiment, one of the questions about your estimates will be chosen for payment. You will receive a bonus of 3 Euros if your guess is close enough to the actual answer. It is in your interest to provide accurate guesses, as this increases the probability of receiving the bonus. If you would like to know more about the mechanism we use to determine whether you receive this bonus, feel free to click on the button below.

Optional: Click here for information about the mechanism

[34] The instructions for the belief elicitation are adapted from Babcock et al. (2017)

[IF CLICKED: If a question is chosen for payment, the probability that you receive a bonus payment of 3 Euros will depend on your prediction error. This prediction error is the distance between your estimate and the correct number. The closer your estimate is to the correct answer, the larger is the probability that you will receive the bonus.

Assume that your actual performance is X solved [summations] [mazes] and you guessed that you had Y solved [summations] [mazes]. In this case your squared prediction error is $(X - Y)^2$. To determine the probability of receiving the bonus, the computer first draws a number between 0 and 40000, let's call this number T. Then this number T is compared to your squared prediction error. If T is larger than the squared error, you will receive the bonus payment for this question. If your squared prediction error is larger than or equal to T, you will not receive a bonus for this question.]

Estimates about performance

You previously estimated the performance of another worker in the [summation task] [maze task]. Now you know your comparable worker's wage.

- **Your estimate was that a worker solved [x] elements correctly.**
- **The wage your comparable worker received in the previous negotiation period was [y] points.**

We would like to know your estimate of your comparable worker's performance. Please enter below how many elements you now think the comparable worker solved correctly.

At the end of the experiment, one of the questions about your estimates will be chosen for payment. You will receive a bonus of 3 Euros if your guess is close enough to the actual answer. It is in your interest to provide accurate guesses, as this increases the probability of receiving the bonus. If you would like to know more about the mechanism we use to determine whether you receive this bonus, feel free to click on the button below.

Optional: Click here for information about the mechanism

[IF CLICKED: If a question is chosen for payment, the probability that you receive a bonus payment of 3 Euros will depend on your prediction error. This prediction error is the distance between your estimate and the correct number. The closer your estimate is to the correct answer, the larger is the probability that you will receive the bonus.

Assume that your actual performance is X solved [summations] [mazes] and you guessed that you had Y solved [summations] [mazes]. In this case your squared prediction error is $(X - Y)^2$. To determine the probability of receiving the bonus, the computer first draws a number between 0 and 20, let's call this number T. Then this number T is compared to your squared prediction error. If T is larger than the squared error, you will receive the bonus payment for this question. If your squared prediction error is larger than or equal to T, you will not receive a bonus for this question.]

CHAPTER

5

SUMMARY

STRATEGIC INTERACTION AND SOCIAL INFORMATION: ESSAYS IN BEHAVIOURAL ECONOMICS

This thesis studies the behavioural economics of strategic interactions and how social information shapes these interactions in distinct settings. Using a broad variety of economic techniques, I consider three types of social information in three types of environments: (1) the ease of implementing selfish actions in groups and its consequences for the group composition, (2) endogenously generated information on the likelihood of success in a project of uncertain quality in team experimentation, and (3) information on others' wages and performance in wage negotiations.

Chapter 2 focuses on the information that defaults in group decisions entail and its influence on group formation. More specifically, we study the effects of varying individual pivotality and the endogeneity of group entry on the selfishness of group decisions. Selfish choices by groups are often linked to the possibility of diffusing responsibility; the moral costs of these decisions appear smaller when individual pivotality is reduced. This chapter investigates whether this characteristic of group decisions attracts particularly selfish individuals and the consequences of self-selection into groups for the outcomes of group decisions. We test this in a laboratory experiment. Our experimental design explores unanimity voting under distinct defaults to identify the effect of diffusion of responsibility, varying whether individuals can self-select into group decision making.

In exogenously formed groups we find evidence of responsibility diffusion, but this diminishes with repetition. In endogenously formed groups, on the other hand, the possibility to diffuse responsibility leads to consistently more selfish choices. Our results demonstrate the role of self-selection in generating differences in group behaviour depending on individual pivotality. Driven by a heterogeneous selection pattern, endogenous group formation amplifies the effects of a change in pivotality. Some people actively seek an environment to diffuse responsibility, while others join groups to promote pro-social behaviour.

Chapter 3 examines both the demand- and supply-side of social information in the context of experimentation, which is at the core of innovation. This chapter studies collaborative experimentation in teams, focusing on the inherent two-dimensional free-riding problem; agents create a payoff and an informational externality, which both induce free riding. This theoretically results in inefficiently low experimentation. In particular, agents' experimentation being observable decreases experimentation because of a discouragement effect. Agents become pessimistic after observing high experimentation but no success. In a laboratory experiment, we study how distinct elements of the experimentation environment affect strategic experimentation. We vary (1) the observability of experimentation, and (2) whether agents work on joint or separate projects.

Teams largely overcome the free-riding problem. Both the observability of experimentation and experimenting jointly increase experimentation levels. There is no lack of sophistication in updating beliefs that drives this, neither do subjects disregard their experimentation's effect on others. Instead, the data can be best explained by joint, observable experimentation creating incentives to 'lead by example' and setting norms of high experimentation.

Chapter 4 concentrates on the role of social information, in particular wage and performance information, in negotiations. Wage transparency regulation is widely considered and adopted as a tool to reduce the gender wage gap. We combine field and laboratory evidence to address how and when wage transparency can be effective and explore the role of correcting beliefs as a mechanism. In the field, this paper studies a German wage transparency policy that allows employees to request wage information of comparable employees. Exploiting variation across firm size and time, we first provide causal evidence that this regulation neither affects wages in general nor the gender wage gap in particular.

In an online laboratory experiment, we study whether the failure of this policy hinges on two aspects: (1) the endogenous availability of wage information, and (2) the absence of performance information. Our data underline the importance of both factors. In contrast to endogenously acquired wage information, exogenously provided wage information does increase overall wages, as does the provision of performance information. However, none of these types of information reduces the gender wage gap. Wage information even deters women from entering negotiations, underlining its limited power in decreasing the gender wage gap.

SAMENVATTING

STRATEGISCHE INTERACTIES EN SOCIALE INFORMATIE: ESSAYS IN GEDRAGSECONOMIE

Dit proefschrift bestudeert de gedragseconomie van strategische interacties en hoe deze interacties in verschillende situaties worden beïnvloed door sociale informatie. Met behulp van een breed scala aan economische technieken analyseer ik de effecten van drie soorten sociale informatie in drie soorten situaties: (1) hoe groepen makkelijker egoïstische beslissingen nemen dan individuen, en de gevolgen daarvan voor de groepssamenstelling, (2) hoe endogeen gegenereerde informatie over de kans op succes van een project van onzekere kwaliteit de inspanning van een onderzoeksteam beïnvloedt, en (3) het effect van informatie over de lonen en prestaties van collega's op loononderhandelingen.

Hoofdstuk 2 onderzoekt de informatiewaarde van standaarden in groepsbeslissingen (de beslissing die wordt uitgevoerd als de groep geen overeenstemming bereikt) en de invloed daarvan op groepsvorming. We bestuderen de effecten van de individuele invloed op de groepsbeslissing en het al of niet zelf kunnen kiezen om in een groep of individueel te beslissen, op het egoïsme van groepsbeslissingen. Het kunnen delen van de verantwoordelijkheid in een groep zorgt voor egoïstischere beslissingen; de morele kosten van deze beslissingen lijken minder wanneer de individuele invloed op de groepsbeslissing kleiner is. Dit hoofdstuk onderzoekt of dit kenmerk van groepsbeslissingen in het bijzonder egoïstische individuen aantrekt, en wat de gevolgen van deze zelfselectie zijn voor de uitkomsten van groepsbeslissingen. We testen dit in een laboratoriumexperiment. Onze experimentele opzet onderzoekt groepsbeslissingen onder unanimiteit onder verschillende standaarden om het effect van gedeelde verantwoordelijkheid te bepalen, afhankelijk van het feit of individuen zelf kunnen kiezen voor groepsbeslissingen.

In exogeen gevormde groepen vinden we aanwijzingen voor gedeelde verantwoordelijkheid, maar dit effect neemt af bij herhaling. In endogeen gevormde groepen (deelnemers hebben ervoor gekozen om in een groep te beslissen) daarentegen leidt de mogelijkheid om verantwoordelijkheid te delen tot consistent meer zelfzuchtige keuzes. Onze resultaten tonen de rol aan van zelfselectie bij het genereren van verschillen in groepsgedrag, afhankelijk van in hoeverre het individu doorslaggevend kan zijn. Gedreven door een heterogeen selectiepatroon versterkt endogene groepsvorming de effecten van een verandering in doorslaggevendheid. Sommige mensen zoeken actief een situatie waarin de verantwoordelijkheid voor egoïstische keuzes met anderen gedeeld wordt, terwijl andere mensen zich juist bij groepen aansluiten om zo pro-sociaal gedrag te bevorderen.

Hoofdstuk 3 onderzoekt de vraag- en aanbodzijde van sociale informatie in de context van onderzoek, de kern van innovatie. Dit hoofdstuk bestudeert samenwerkende onderzoeksteams en richt zich op het inherente tweedimensionale meeliftprobleem met zowel een belonings- als een informatieve externaliteit. Theoretisch zou dit moeten leiden tot minder onderzoek dan efficiënt is. Wanneer onderzoeksinspanningen van anderen waarneembaar zijn, veroorzaakt dit een ontmoedigingseffect. Onderzoekers worden pessimistisch na het waarnemen van veel inspanningen zonder doorbraak. In een laboratoriumexperiment bestudeer ik hoe verschillende elementen van de situatie strategische onderzoeksinspanningen beïnvloeden. Ik varieer (1) de observeerbaarheid van de onderzoeksinspanning, en (2) of onderzoekers aan gezamenlijke of afzonderlijke projecten werken.

Teams ondervangen grotendeels het meeliftprobleem. Zowel de observeerbaarheid van de onderzoeksinspanningen als het gezamenlijk onderzoeken verhogen de inspanningen. Dit wordt niet veroorzaakt doordat verwachtingen niet juist worden aangepast, en ook niet doordat deelnemers het effect van hun inspanningen op de verwachtingen van anderen negeren. De resultaten kunnen het best worden verklaard doordat gezamenlijke, waarneembare onderzoeksinspanningen stimuleren om "het goede voorbeeld te geven" en om zo sociale normen te beïnvloeden.

Hoofdstuk 4 concentreert zich op de rol van sociale informatie, met name loon- en prestatie-informatie van collega's, bij loononderhandelingen. Om de loonkloof tussen mannen en vrouwen te verkleinen wordt in het algemeen regelgeving over loontransparantie aanbevolen. Wij combineren gegevens uit de praktijk en het laboratorium om na te gaan hoe en wanneer loontransparantie effectief kan zijn. In de praktijk bestuderen we een Duitse regelgeving over loontransparantie waarbij werknemers looninformatie van vergelijkbare werknemers kunnen opvragen. Door gebruik te maken van variatie in bedrijfsgrootte en tijd, bewijzen we dat deze regeling noch de lonen in het algemeen, noch de loonkloof tussen mannen en vrouwen in het bijzonder beïnvloedt.

In een online laboratoriumexperiment onderzoeken wij of het mislukken van dit beleid afhangt van twee aspecten: (1) de endogene beschikbaarheid van looninformatie, en (2) de afwezigheid van prestatie-informatie. Onze analyse benadrukt het belang van beide factoren. In tegenstelling tot endogeen verkregen looninformatie, verhoogt exogeen verstrekte looninformatie wel de totale lonen. Hetzelfde geldt voor prestatie- informatie. Geen van deze soorten informatie vermindert echter de loonkloof tussen mannen en vrouwen. Looninformatie weerhoudt vrouwen er zelfs van om aan onderhandelingen deel te nemen, wat de beperkte rol van transparantie bij het verkleinen van de loonkloof tussen mannen en vrouwen benadrukt.

BIBLIOGRAPHY

Ahn, T.-K., R. M. Isaac, and T. C. Salmon (2008): "Endogenous group formation," *Journal of Public Economic Theory*, 10, 171–194.

Allcott, H., and T. Rogers (2014): "The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation," *American Economic Review*, 104, 3003–37.

Almås, I., A. W. Cappelen, and B. Tungodden (2020): "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" *Journal of Political Economy*, 128, 1753–1788.

Amanatullah, E. T., and C. H. Tinsley (2013): "Punishing female negotiators for asserting too much... or not enough: Exploring why advocacy moderates backlash against assertive female negotiators," *Organizational Behavior and Human Decision Processes*, 120, 110–122.

Anderson, C. M. (2012): "Ambiguity aversion in multi-armed bandit problems," *Theory and Decision*, 72, 15–33.

Anderson, L. R., and C. A. Holt (1997): "Information cascades in the laboratory," *American Economic Review*, 87, 847–862.

Andreoni, J., and R. Petrie (2004): "Public goods experiments without confidentiality: A glimpse into fund-raising," *Journal of Public Economics*, 88, 1605–1623.

Azmat, G., and B. Petrongolo (2014): "Gender and the labor market: What have we learned from field and lab experiments?" *Labour Economics*, 30, 32–40.

Babcock, L., M. P. Recalde, L. Vesterlund, and L. Weingart (2017): "Gender differences in accepting and receiving requests for tasks with low promotability," *American Economic Review*, 107, 714–47.

Baker, M., Y. Halberstam, K. Kroft, A. Mas, and D. Messacar (forthcoming): "Pay Transparency and the Gender Gap," *American Economic Journal: Applied Economics*.

Bamieh, O., and L. Ziegler (2022): "Can Wage Transparency Alleviate Gender Sorting in the Labor Market?" IZA Discussion Paper 15363, IZA Institute of Labor Economics.

Banholzer, M., F. Metzeler, and E. Roth (2019): "Fielding high-performing innovation teams," Strategy & Corporate Finance Report, McKinsey & Company.

Banks, J., M. Olson, and D. Porter (1997): "An experimental analysis of the bandit problem," *Economic Theory*, 10, 55–77.

Bartling, B., and U. Fischbacher (2012): "Shifting the blame: On delegation and responsibility," *Review of Economic Studies*, 79, 67–87.

Bartling, B., U. Fischbacher, and S. Schudy (2015): "Pivotality and responsibility attribution in sequential voting," *Journal of Public Economics*, 128, 133–139.

Behnk, S., L. Hao, and E. Reuben (2022): "Shifting normative beliefs: On why groups behave more antisocially than individuals," *European Economic Review*, 145, 104116.

Benjamin, D. J. (2019): "Errors in probabilistic reasoning and judgment biases," *Handbook of Behavioral Economics: Applications and Foundations 1*, 2, 69–186.

Bennedsen, M., E. Simintzi, M. Tsoutsoura, and D. Wolfenzon (2022): "Do firms respond to gender pay gap transparency?" *The Journal of Finance*, 77, 2051–2091.

Biasi, B., and H. Sarsons (2022): "Flexible wages, bargaining, and the gender gap," *The Quarterly Journal of Economics*, 137, 215–266.

Black, J., N. Hashimzade, and G. Myles (2009): *A Dictionary of Economics*: Oxford University Press.

Blau, F. D., and L. M. Kahn (2017): "The gender wage gap: Extent, trends, and explanations," *Journal of Economic Literature*, 55, 789–865.

Blundell, J. (2021): "UK gender pay gap reporting: a crude but effective policy?" LSE busines review, London School of Economics and Political Science.

Böheim, R., and S. Gust (2021): "The Austrian pay transparency law and the gender wage gap," IZA Discussion Paper 14206, IZA Institute of Labor Economics.

Bolton, P., and C. Harris (1999): "Strategic experimentation," *Econometrica*, 67, 349–374.

Bonatti, A., and J. Hörner (2011): "Collaborating," *American Economic Review*, 101, 632–663.

Boneva, T., T. Buser, A. Falk, and F. Kosse (2022): "The origins of gender differences in competitiveness and earnings expectations: Causal evidence from a mentoring intervention," CEPR Discussion Paper DP17008, Centre for Economic Policy Research.

Bornstein, G., and I. Yaniv (1998): "Individual and group behavior in the ultimatum game: Are groups more "rational" players?" *Experimental Economics*, 1, 101–108.

Bouckaert, S., A. F. Pales, C. McGlade, U. Remme, B. Wanner, L. Varro, D. D'Ambrosio, and T. Spencer (2021): "Net Zero by 2050: A Roadmap for the Global Energy Sector," Flagship report, International Energy Agency.

Bowles, H. R., L. Babcock, and K. L. McGinn (2005): "Constraints and triggers: situational mechanics of gender in negotiation.," *Journal of Personality and Social Psychology*, 89, 951.

Boyce, J. R., D. M. Bruner, and M. McKee (2016): "Strategic experimentation in the lab," *Managerial and Decision Economics*, 37, 375–391.

Brandts, J., K. Gërxhani, and A. Schram (2020): "Are there gender differences in status-ranking aversion?" *Journal of Behavioral and Experimental Economics*, 84, 101485.

Brekke, K. A., K. E. Hauge, J. T. Lind, and K. Nyborg (2011): "Playing with the good guys. A public good game with endogenous group formation," *Journal of Public Economics*, 95, 1111–1118.

Briel, S., A. Osikominu, G. Pfeifer, M. Reutter, and S. Satlukal (2022): "Gender differences in wage expectations: the role of biased beliefs," *Empirical Economics*, 62, 187–212.

Brütt, K. (2020): "Collaborating in strategic experimentation," Pre-registration 5503, AEA RCT Registry.

Brütt, K., and H. Yuan (2021): "Can transparency alleviate the gender pay gap?" Pre-registration 7593, AEA RCT Registry.

——— (2022): "Pitfalls of pay transparency: Evidence from the lab and the field," Tinbergen Institute Discussion Paper 2022-055/I, Tinbergen Institute.

Brütt, K., A. Schram, and J. Sonnemans (2020): "Endogenous group formation and responsibility diffusion: An experimental study," *Games and Economic Behavior*, 121, 1–31.

Calonico, S., M. D. Cattaneo, and M. H. Farrell (2020): "Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs," *The Econometrics Journal*, 23, 192–210.

Cameron, A. C., and D. L. Miller (2015): "A practitioner's guide to cluster-robust inference," *Journal of Human Resources*, 50, 317–372.

Cappelen, A. W., A. D. Hole, E. Ø. Sørensen, and B. Tungodden (2007): "The pluralism of fairness ideals: An experimental approach," *American Economic Review*, 97, 818–827.

Card, D., A. Mas, E. Moretti, and E. Saez (2012): "Inequality at work: The effect of peer salaries on job satisfaction," *American Economic Review*, 102, 2981–3003.

Cason, T. N., and V. Mui (1997): "A laboratory study of group polarisation in the team dictator game," *The Economic Journal*, 107, 1465–1483.

Charness, G. (2000): "Responsibility and effort in an experimental labor market," *Journal of Economic Behavior & Organization*, 42, 375–384.

——— (2004): "Attribution and reciprocity in an experimental labor market," *Journal of Labor Economics*, 22, 665–688.

Charness, G., T. Garcia, T. Offerman, and M. C. Villeval (2020): "Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory?" *Journal of Risk and Uncertainty*, 60, 99–123.

Charness, G., and M. Rabin (2002): "Understanding social preferences with simple tests," *The Quarterly Journal of Economics*, 117, 817–869.

Charness, G., L. Rigotti, and A. Rustichini (2007): "Individual behavior and group membership," *American Economic Review*, 97, 1340–1352.

Charness, G., and M. Sutter (2012): "Groups make better self-interested decisions," *Journal of Economic Perspectives*, 26, 157–176.

Chen, R. (2017): "Coordination with endogenous groups," *Journal of Economic Behavior & Organization*, 141, 177–187.

Chen, Y., F. M. Harper, J. Konstan, and S. X. Li (2010): "Social comparisons and contributions to online communities: A field experiment on movielens," *American Economic Review*, 100, 1358–98.

Chen, Y., and S. X. Li (2009): "Group identity and social preferences," *American Economic Review*, 99, 431–457.

Coffman, L. C., C. R. Featherstone, and J. B. Kessler (2017): "Can social information affect what job you choose and keep?" *American Economic Journal: Applied Economics*, 9, 96–117.

Crawford, V. (1997): "Theory and experiment in the analysis of strategic interaction," in *Behavioral Game Theory* ed. by Camerer, C.: Princeton, NJ: Princeton University Press.

Croson, R., E. Fatas, and T. Neugebauer (2005): "Reciprocity, matching and conditional cooperation in two public goods games," *Economics Letters*, 87, 95–101.

Crucian, G. P., and S. A. Berenbaum (1998): "Sex differences in right hemisphere tasks," *Brain and Cognition*, 36, 377–389.

Cullen, Z. B., and B. Pakzad-Hurson (2021): "Equilibrium effects of pay transparency," NBER Working paper 28903, National Bureau of Economic Research.

Cullen, Z., and R. Perez-Truglia (2022): "How much does your boss make? The effects of salary comparisons," *Journal of Political Economy*, 130, 766–822.

Dana, J., D. M. Cain, and R. M. Dawes (2006): "What you don't know won't hurt me: Costly (but quiet) exit in dictator games," *Organizational Behavior and Human Decision Processes*, 100, 193–201.

Dana, J., R. A. Weber, and J. X. Kuang (2007): "Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness," *Economic Theory*, 33, 67–80.

Danz, D., L. Vesterlund, and A. J. Wilson (2022): "Belief Elicitation and Behavioral Incentive Compatibility," *American Economic Review*, 9.

Darley, J. M., and B. Latané (1968): "Bystander intervention in emergencies: Diffusion of responsibility.," *Journal of Personality and Social Psychology*, 8, 377–383.

Dong, Y., H. Ma, Z. Shen, and K. Wang (2017): "A century of science: Globalization of scientific collaborations, citations, and innovations," in *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, 1437–1446.

Duchini, E., S. Simion, and A. Turrell (2020): “Pay transparency and cracks in the glass ceiling,” CAGE working paper 482, Centre for Competitive Advantage in the Global Economy.

Ehrhart, K.-M., and C. Keser (1999): “Mobility and cooperation: On the run,” CIRANO working paper 99s-24, Centre interuniversitaire de recherche en analyse des organisations.

Ellguth, P., and S. Kohaut (2019): “A note on the decline of collective bargaining coverage: The role of structural change,” *Jahrbücher für Nationalökonomie und Statistik*, 239, 39–66.

Engl, F. (2018): “A Theory of Causal Responsibility Attribution,” CESifo Working Paper 9898, Center for Economic Studies and ifo Institute.

Erkal, N., L. Gangadharan, and B. H. Koh (2020): “Replication: Belief Elicitation with Quadratic and Binarized Scoring Rules,” *Journal of Economic Psychology*, 81, 102315.

von Essen, E., M. Huysentruyt, and T. Miettinen (2020): “Exploration in teams and the encouragement effect: Theory and experimental evidence,” *Management Science*, 66, 5861–5885.

European Commission (2021): “Pay Transparency: Commission proposes measures to ensure equal pay for equal work,” Press release, European Commission.

Eurostat (2021): “Gender pay gap in unadjusted form,” data retrieved from `https://ec.europa.eu/eurostat/databrowser/view/sdg_05_20/default/table?lang=en`.

Exley, C. L., M. Niederle, and L. Vesterlund (2020): “Knowing when to ask: The cost of leaning in,” *Journal of Political Economy*, 128, 816–854.

Falk, A., T. Neuber, and N. Szech (2020): “Diffusion of being pivotal and immoral outcomes,” *The Review of Economic Studies*, 87, 2205–2229.

Fehr, E., and K. M. Schmidt (1999): “A theory of fairness, competition, and cooperation,” *Quarterly Journal of Economics*, 114, 817–868.

Fehr, E., and T. Williams (2018): “Social norms, endogenous sorting and the culture of cooperation,” IZA Discussion Paper 11457, IZA Institute of Labor Economics.

Fischbacher, U., S. Gächter, and E. Fehr (2001): “Are people conditionally cooperative? Evidence from a public goods experiment,” *Economics Letters*, 71, 397–404.

GAMAGE, D. D. K., G. KAVETSOS, S. MALLICK, AND A. SEVILLA (2020): "Pay Transparency Initiative and Gender Pay Gap: Evidence from Research-Intensive Universities in the UK," IZA Discussion Paper 13635, IZA Institute of Labor Economics.

GIHLEB, R., R. LANDSMAN, AND L. VESTERLUND (2020): "The Effect of Negotiation on Securing Equal Pay for Equal Work," Working paper.

GNEEZY, U., M. NIEDERLE, AND A. RUSTICHINI (2003): "Performance in competitive environments: Gender differences," *The Quarterly Journal of Economics*, 118, 1049–1074.

GREIG, F. (2008): "Propensity to negotiate and career advancement: Evidence from an investment bank that women are on a "slow elevator"," *Negotiation Journal*, 24, 495–508.

GREMBI, V., T. NANNICINI, AND U. TROIANO (2016): "Do fiscal rules matter?" *American Economic Journal: Applied Economics*, 8, 1–30.

GROSSMAN, Z., AND J. J. VAN DER WEELE (2017): "Self-image and willful ignorance in social decisions," *Journal of the European Economic Association*, 15, 173–217.

GULYAS, A., S. SEITZ, AND S. SINHA (forthcoming): "Does pay transparency affect the gender wage gap? Evidence from Austria," *American Economic Journal: Economic Policy*.

GÜRERK, Ö., B. IRLENBUSCH, AND B. ROCKENBACH (2006): "The competitive advantage of sanctioning institutions," *Science*, 312, 108–111.

——— (2014): "On cooperation in open communities," *Journal of Public Economics*, 120, 220–230.

GÜTH, W., M. V. LEVATI, M. SUTTER, AND E. VAN DER HEIJDEN (2007): "Leading by example with and without exclusion power in voluntary contribution experiments," *Journal of Public Economics*, 91, 1023–1042.

GÜTH, W., R. SCHMITTBERGER, AND B. SCHWARZE (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3, 367–388.

HAECKL, S. (2022): "Image concerns in ex-ante self-assessments–Gender differences and behavioral consequences," *Labour Economics*, 76, 102166.

HAMILTON, B. H., J. A. NICKERSON, AND H. OWAN (2003): "Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation," *Journal of Political Economy*, 111, 465–497.

Harrison, G. W., J. Martínez-Correa, and J. T. Swarthout (2014): "Eliciting subjective probabilities with binary lotteries," *Journal of Economic Behavior & Organization*, 101, 128–140.

Herbst, L., K. A. Konrad, and F. Morath (2015): "Endogenous group formation in experimental contests," *European Economic Review*, 74, 163–189.

Hernandez-Arenaz, I., and N. Iriberri (2018): "Women ask for less (only from men): Evidence from bargaining in the field," *Journal of Economic Behavior & Organization*, 152, 192–214.

Hett, F., M. Mechtel, and M. Kröll (2020): "The structure and behavioral effects of revealed social identity preferences," *The Economic Journal*, 130, 2569–2595.

Hoelzemann, J., and N. Klein (2021): "Bandits in the Lab," *Quantitative Economics*, 12, 1021–1051.

Holmstrom, B. (1982): "Moral hazard in teams," *The Bell Journal of Economics*, 13, 324–340.

Holt, C. A., and S. K. Laury (2002): "Risk aversion and incentive effects," *American Economic Review*, 92, 1644–1655.

Hörner, J., and A. Skrzypacz (2017): "Learning, experimentation and information design," *Advances in Economics and Econometrics*, 1, 63–98.

Hossain, T., and R. Okui (2013): "The binarized scoring rule," *Review of Economic Studies*, 80, 984–1001.

Hudja, S., and D. Woods (2021): "Exploration Versus Exploitation: A Laboratory Test of the Single-Agent Exponential Bandit Model," Available at SSRN 484498.

Irlenbusch, B., and D. Saxler (2019): "The role of social information, market framing, and diffusion of responsibility as determinants of socially responsible behavior," *Journal of Behavioral and Experimental Economics*, 80, 141–161.

Kaiser, J. (2007): "An exact and a Monte Carlo proposal to the Fisher–Pitman permutation tests for paired replicates and for independent samples," *The Stata Journal*, 7, 402–412.

Keller, G., and S. Rady (2010): "Strategic experimentation with Poisson bandits," *Theoretical Economics*, 5, 275–311.

——— (2015): "Breakdowns," *Theoretical Economics*, 10, 175–202.

Keller, G., S. Rady, and M. Cripps (2005): "Strategic experimentation with exponential bandits," *Econometrica*, 73, 39–68.

Kiessling, L., P. Pinger, P. Seegers, and J. Bergerhoff (2019): "Gender Differences in Wage Expectations: Sorting, Children, and Negotiation Styles," IZA Discussion Paper 12522, IZA Institute of Labor Economics.

Kline, P., and A. Santos (2012): "A score based approach to wild bootstrap inference," *Journal of Econometric Methods*, 1, 23–41.

Kocher, M. G., T. Cherry, S. Kroll, R. J. Netzer, and M. Sutter (2008): "Conditional cooperation on three continents," *Economics Letters*, 101, 175–178.

Kocher, M. G., S. Schudy, and L. Spantig (2017): "I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups," *Management Science*, 1–14.

Kocher, M., S. Strau, and M. Sutter (2006): "Individual or team decision-making—causes and consequences of self-selection," *Games and Economic Behavior*, 56, 259–270.

Konow, J. (2000): "Fair shares: Accountability and cognitive dissonance in allocation decisions," *American Economic Review*, 90, 1072–1091.

Kosfeld, M., A. Okada, and A. Riedl (2009): "Institution formation in public goods games," *American Economic Review*, 99, 1335–1355.

Kretchmer, H. (2020): "From dining pods to see-through masks: 6 ways innovations are helping in the pandemic," Online article, World Economic Forum.

Kugler, T., E. E. Kausel, and M. G. Kocher (2012): "Are groups more rational than individuals? A review of interactive decision making in groups," *Wiley Interdisciplinary Reviews: Cognitive Science*, 3, 471–482.

Kwon, O. (2020): "Strategic Experimentation with Uniform Bandit: An Experimental Study," Working paper.

Latané, B., and J. M. Darley (1968): "Group inhibition of bystander intervention in emergencies.," *Journal of Personality and Social Psychology*, 10, 215–221.

Lazear, E. P., U. Malmendier, and R. A. Weber (2012): "Sorting in Experiments with Application to Social Preferences," *American Economic Journal: Applied Economics*, 4, 136–163.

Leibbrandt, A., and J. A. List (2015): "Do women avoid salary negotiations? Evidence from a large-scale natural field experiment," *Management Science*, 61, 2016–2024.

LEVATI, M. V., M. SUTTER, AND E. VAN DER HEIJDEN (2007): "Leading by example in a public goods experiment with heterogeneity and incomplete information," *Journal of Conflict Resolution*, 51, 793–818.

LIEBRAND, W. B. (1984): "The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game," *European Journal of Social Psychology*, 14, 239–264.

LUHAN, W. J., M. G. KOCHER, AND M. SUTTER (2009): "Group polarization in the team dictator game reconsidered," *Experimental Economics*, 12, 26–41.

MAS, A. (2017): "Does transparency lead to pay compression?" *Journal of Political Economy*, 125, 1683–1721.

MAZEI, J., J. HÜFFMEIER, P. A. FREUND, A. F. STUHLMACHER, L. BILKE, AND G. HERTEL (2015): "A meta-analysis on gender differences in negotiation outcomes and their moderators.," *Psychological Bulletin*, 141, 85.

MEIDINGER, C., AND M. C. VILLEVAL (2002): "Leadership in teams: signaling or reciprocating?" Working paper 02-13, GATE.

MEYER, R. J., AND Y. SHI (1995): "Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem," *Management Science*, 41, 817–834.

MOIR, R. (1998): "A Monte Carlo analysis of the Fisher randomization technique: reviving randomization for experimental economists," *Experimental Economics*, 1, 87–100.

MOORE, D. A., AND P. J. HEALY (2008): "The trouble with overconfidence.," *Psychological Review*, 115, 502.

MOSCOVICI, S., AND M. ZAVALLONI (1969): "The group as a polarizer of attitudes.," *Journal of Personality and Social Psychology*, 12, 125–135.

NICKLISCH, A., K. GRECHENIG, AND C. THÖNI (2016): "Information-sensitive leviathans," *Journal of Public Economics*, 144, 1–13.

NIEDERLE, M., AND L. VESTERLUND (2007): "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics*, 122, 1067–1101.

OLDEN, A., AND J. MØEN (2022): "The triple difference estimator," *The Econometrics Journal*, 25, 531–553.

POTTERS, J., M. SEFTON, AND L. VESTERLUND (2005): "After you—endogenous sequencing in voluntary contribution games," *Journal of Public Economics*, 89, 1399–1419.

——— (2007): "Leading-by-example and signaling in voluntary contribution games: An experimental study," *Economic Theory*, 33, 169–182.

Recalde, M. P., and L. Vesterlund (2022): "Gender Differences in Negotiation and Policy for Equalizing Outcomes," in *Bargaining*: Springer, 455–475.

Riedl, A., I. M. Rohde, and M. Strobel (2015): "Efficient coordination in weakest-link games," *The Review of Economic Studies*, 83, 737–767.

Rigdon, M. L. (2012): "An experimental investigation of gender differences in wage negotiations," Available at SSRN 2165253.

Robbett, A. (2014): "Local institutions and the dynamics of community sorting," *American Economic Journal: Microeconomics*, 6, 136–156.

Rothenhäusler, D., N. Schweizer, and N. Szech (2018): "Guilt in voting and public good games," *European Economic Review*, 101, 664–681.

Roussille, N. (2020): "The central role of the ask gap in gender pay inequality," Working paper, UC Berkeley.

Ruf, K., L. Schmidtlein, S. Seth, H. Stüber, M. Umkehrer et al. (2021): "Linked Employer-Employee Data from the IAB: LIAB Longitudinal Model (LIAB LM) 1975-2019," Data report, Institut für Arbeitsmarkt-und Berufsforschung (IAB).

Sanchis-Segura, C., N. Aguirre, Á. J. Cruz-Gómez, N. Solozano, and C. Forn (2018): "Do gender-related stereotypes affect spatial performance? Exploring when, how and to whom using a chronometric two-choice mental rotation task," *Frontiers in Psychology*, 9, 1261.

Säve-Söderbergh, J. (2019): "Gender gaps in salary negotiations: Salary requests and starting salaries in the field," *Journal of Economic Behavior & Organization*, 161, 35–51.

Schlag, K. H., J. Tremewan, and J. J. Van der Weele (2015): "A penny for your thoughts: A survey of methods for eliciting beliefs," *Experimental Economics*, 18, 457–490.

Schmidtlein, L., S. Seth, P. Vom Berge et al. (2020): "Stichprobe Integrierter Employer Employee Daten (SIEED) 1975 2018," Data report, Institut für Arbeitsmarkt- und Berufsforschung (IAB).

Schopler, J., C. A. Insko, S. M. Drigotas, J. Wieselquist, M. B. Pemberton, and C. Cox (1995): "The role of identifiability in the reduction of interindividual-intergroup discontinuity," *Journal of Experimental Social Psychology*, 31, 553–574.

SCHRAM, A., J. BRANDTS, AND K. GËRXHANI (2019): "Social-status ranking: a hidden channel to gender inequality under competition," *Experimental Economics*, 22, 396–418.

SHANG, J., AND R. CROSON (2009): "A field experiment in charitable contribution: The impact of social information on the voluntary provision of public goods," *The Economic Journal*, 119, 1422–1439.

SIEGEL, S., AND N. J. CASTELLAN (1981): *Nonparametric Statistics for the Behavioral Sciences*, New York: McGraw-Hill.

SUGDEN, R. (1984): "Reciprocity: the supply of public goods through voluntary contributions," *The Economic Journal*, 94, 772–787.

SUTTER, M. (2009): "Individual behavior and group membership: Comment," *American Economic Review*, 99, 2247–2257.

SUTTER, M., S. HAIGNER, AND M. G. KOCHER (2010): "Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations," *The Review of Economic Studies*, 77, 1540–1566.

TAJFEL, H. (1974): "Social identity and intergroup behaviour," *Social Science Information*, 13, 65–93.

THOMKE, S. H. (2003): *Experimentation matters: Unlocking the potential of new technologies for innovation*: Harvard Business Press.

THÖNI, C., AND S. VOLK (2018): "Conditional cooperation: Review and refinement," *Economics Letters*, 171, 37–40.

VESTERLUND, L. (2003): "The informational value of sequential fundraising," *Journal of Public Economics*, 87, 627–657.

WEBER, M., AND A. SCHRAM (2017): "The non-equivalence of labour market taxes: A real-effort experiment," *The Economic Journal*, 127, 2187–2215.

WEIZSÄCKER, G. (2010): "Do we follow others when we should? A simple test of rational expectations," *American Economic Review*, 100, 2340–60.

WERNER, P. (2019): "Wage negotiations and strategic responses to transparency," Working paper, Maastricht University.

WONG, C. H., K. W. SIAH, AND A. W. LO (2019): "Estimation of clinical trial success rates and related parameters," *Biostatistics*, 20, 273–286.

WUCHTY, S., B. F. JONES, AND B. UZZI (2007): "The increasing dominance of teams in production of knowledge," *Science*, 316, 1036–1039.

List of co-authors and contributions

Chapter 3 of this thesis is single-authored, Chapter 2 and Chapter 4 are based on co-authored work. All references of co-authored work are provided in the chapters. The contributions of the individual authors in the co-authored chapters are outlined below.

Chapter 2: "Endogenous group formation and responsibility diffusion: An experimental study"
Co-authors: Arthur Schram and Joep Sonnemans

This chapter builds on Katharina Brütt's MPhil thesis. Katharina Brütt proposed the initial research idea, developed the (theoretical) hypotheses to be tested in the experiment, conducted the analysis of the experimental data, and wrote the first draft of the working paper. All authors jointly developed the experimental design and improved and revised the working paper.

Chapter 4: "Pitfalls of pay transparency: Evidence from the lab and the field."
Co-author: Huaiping Yuan

Katharina Brütt proposed the initial research idea and wrote large parts of the first draft of the working paper. Both authors jointly developed the empirical strategy to analyse the administrative data, formed the hypotheses to be tested in the experiment, designed the experiment, conducted the econometric analysis of the administrative and experimental data, and improved and revised the working paper.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. For a full list of PhD theses that appeared in the series we refer to List of PhD Theses – Tinbergen.nl. The following books recently appeared in the Tinbergen Institute Research Series:

761. M.A. Cotofan, *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*
762. B.P.J. Andrée, *Theory and Application of Dynamic Spatial Time Series Models*
763. P. Pelzl, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*
764. D.M. Kunst, *Essays on Technological Change, Skill Premia and Development*
765. A.J. Hummel, *Tax Policy in Imperfect Labor Markets*
766. T. Klein, *Essays in Competition Economics*
767. M. Vigh, *Climbing the Socioeconomic Ladder: Essays on Sanitation and Schooling*
768. Y. Xu, *Eliciting Preferences and Private Information: Tell Me What You Like and What You Think*
769. S. Rellstab, *Balancing Paid Work and Unpaid Care over the Life-Cycle*
770. Z. Deng, *Empirical Studies in Health and Development Economics*
771. L. Kong, *Identification Robust Testing in Linear Factor Models*
772. I. Neamţu, *Unintended Consequences of Post-Crisis Banking Reforms*
773. B. Klein Teeselink, *From Mice to Men: Field Studies in Behavioral Economics*
774. B. Tereick, *Making Crowds Wiser: The Role of Incentives, Individual Biases, and Improved Aggregation*
775. A. Castelein, *Models for Individual Responses*
776. D. Kolesnyk, *Consumer Disclosures on Social Media Platforms: A Global Investigation*
777. M.A. Rola-Janicka, *Essays on Financial Instability and Political Economy of Regulation*
778. J.J. Klingen, *Natural Experiments in Environmental and Transport Economics*
779. E.M. Aermann, *Educational Choices and Family Outcomes*
780. F.J. Ostermeijer, *Economic Analyses of Cars in the City*
781. T. Özden, *Adaptive Learning and Monetary Policy in DSGE Models*
782. D. Wang, *Empirical Studies in Financial Stability and Natural Capital*
783. L.S. Stephan, *Estimating Diffusion and Adoption Parameters in Networks New Estimation Approaches for the Latent-Diffusion-Observed-Adoption Model*
784. S.R. Mayer, *Essays in Financial Economics*

785. A.R.S. Woerner, *Behavioral and Financial Change – Essays in Market Design*
786. M. Wiegand, *Essays in Development Economics*
787. L.M. Treuren, *Essays in Industrial Economics - Labor market imperfections, cartel stability, and public interest cartels*
788. D.K. Brands, *Economic Policies and Mobility Behaviour*
789. H.T.T. Nguyen, *Words Matter? Gender Disparities in Speeches, Evaluation and Competitive Performance*
790. C.A.P Burik, *The Genetic Lottery. Essays on Genetics, Income, and Inequality*
791. S.W.J. Olijslagers, *The Economics of Climate Change: on the Role of Risk and Preferences*
792. C.W.A. van der Kraats, *On Inequalities in Well-Being and Human Capital Formation*
793. Y. Yue, *Essays on Risk and Econometrics*
794. E.F. Janssens, *Estimation and Identification in Macroeconomic Models with Incomplete Markets*
795. P.B. Kastelein, *Essays in Household Finance: Pension Funding, Housing and Consumption over the Life Cycle*
796. J.O. Oorschot, *Extremes in Statistics and Econometrics*
797. S.D.T. Hoey, *Economics on Ice: Research on Peer Effects, Rehiring Decisions and Worker Absenteeism*
798. J. Vidiella-Martin, *Levelling the Playing Field: Inequalities in early life conditions and policy responses*
799. Y. Xiao, *Fertility, parental investments and intergenerational mobility*
800. X. Yu, *Decision Making under Different Circumstances: Uncertainty, Urgency, and Health Threat*
801. G. Gianluca, *Productivity and Strategies of Multiproduct Firms*
802. H. Kweon, *Biological Perspective of Socioeconomic Inequality*
803. D.K. Dimitrov, *Three Essays on the Optimal Allocation of Risk with Illiquidity, Intergenerational Sharing and Systemic Institutions*
804. J.B. Bloomfield, *Essays on Early Childhood Interventions*
805. S. Yu, *Trading and Clearing in Fast-Paced Markets*
806. M.G. Gregori, *Advanced Measurement and Sampling for Marketing Research*
807. O.C. Soons, *The Past, Present, and Future of the Euro Area*
808. D. Garces Urzainqui, *The Distribution of Development. Essays on Economic Mobility, Inequality and Social Change*
809. A.C. Peker, *Guess What I Think: Essays on the Wisdom in Meta-predictions*
810. A. Akdeniz, *On the Origins of Human Sociality*