Destructive Behavior in Competitive Settings: Motives and Mitigation Mechanisms

Inauguraldissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaften
der Universität Mannheim

vorgelegt von

Jonathan Stäbler

im Frühjahrs-/Sommersemester 2024

Abteilungssprecher:	Prof. Klaus Adam, Ph.D.
Referent:	Prof. Dr. Henrik Orzen
Koreferent:	Prof. Dr. Wladislaw Mill
Vorsitzender der Disputation:	Prof. Dr. Eckhard Janeba
Tag der Disputation:	15. Juli 2024

Acknowledgements

First of all, I want to thank my advisors, Henrik Orzen and Wladislaw Mill, for their excellent supervision and for always being there for me throughout my entire Ph.D. journey. I cannot thank you enough for everything I learned from you and for the time and effort you invested in my academic and personal growth.

I remember numerous very long meetings with Henrik, where we intensively discussed my research. I am profoundly grateful for your meticulously detailed feedback at every stage of my projects. I especially appreciate you teaching me how to tell an honest story and how to be concise, engaging, and courageous. To me, you are a role model for combining academic excellence with authenticity, kindness, and integrity.

I am deeply grateful for all the support I received from Wlada and that I could always walk into your office to ask even small questions. I want to thank you for your valuable insights on research and how to work conferences, how to handle submissions (and rejections), and for pushing me to reach out to relevant academics in the field. Our usual 5-minute chats in the mornings have always set a positive start for my work days. Thank you for everything!

I feel very honored to be a part of the University of Mannheim, where I have enjoyed a pleasant and productive atmosphere. My gratitude goes out to everyone who has provided me with feedback and guidance, and especially to Konstantin Chatziathanasiou, Antonio Ciccone, Peter Duersch, Adrian Hillenbrand, Martin Kesternich, and Cornelius Schneider. I also thank the Graduate School of Economic and Social Sciences at the University of Mannheim for providing financial support and a research-focused environment.

My research also greatly benefited from participation at different summer schools and conferences and I am grateful to everyone I met there for inspiring conversations, feedback, and great experiences. I especially want to thank Subhasish Chowdhury for your invaluable feedback and support! I also want to thank my coauthors, Katarína Čellárová, Raphael Epperson, and Wlada. I am very grateful for everything that I could learn from you and the fun that we are having doing research together. It is great that not only we are colleagues, who have fruitful work-related discussions but also that I can call you my friends.

I was very lucky to have a supportive cohort at CDSE, with whom I had a lot of fun, intensive study sessions, and passionate and heated discussions about many topics related to life, research, and academia. Thank you to Giovanni Ballarin, Suzanne Bellue, Jacopo Gambato, Tommaso Gasparini, Boris Ginzburg, Aleksandra Khokhlova, Cenchen Liu, Lukas Mahler, Valentina Melentyeva, Laura Montenbruck, Andrés Plúas Lopez, Oliver Pfäuti, Jasmina Simon, Fabian Seyrich, and Samuel Siewers.

For me, six years of Ph.D. meant having six different offices, so I want to thank all my office buddies and especially Oliver and Mark, who gave me all kinds of support I needed and lightened up my days. My special thanks also go to Argun Aman and Michael Hilweg, with whom I haven't shared an office yet but had many nourishing and deep conversations. Thank you for our friendship and support!

Apparently, six years of Ph.D. also means receiving administrative support from six different CDSE managers. I am grateful to Dagmar Röttsches, Golareh Khalilpour, Caroline Mohr, Ulrich Kehl, Lars Biesewig, and Kristina Kadel. Likewise, I am grateful to Friederike Pipphardt, Sylvia Rosenkranz, and Marion Lehnert, who supported me with different parts of academic administration and I received great IT support from Thomas Behles and Patrick Schmitt. I also appreciate the help from the former mLab manager Franziska Heinicke for running experiments.

I want to thank my friends outside of academia, especially Jonas, Lorenz, Matthias, Lingo, Rüdi, and Kneller for being there for me, for your refreshing perspectives, and for our memorable adventures together. Likewise, I am grateful to my roommates during the first years of my time in Mannheim – Lisa, Svenja, and Pierre – for providing a safe haven and a feeling of home. I will always remember our joyful time together at Jungbuschstraße.

I am deeply grateful to my family, who always supported me. I want to thank my parents, Bernadette & Dieter, for your immense love. You provided me with a deep sense of safety and trust, nourished optimism, and made me feel deserving. Thank you for always wanting and pushing for the best for me. I am very grateful to my brothers Thomas, Dominik, and Samuel, who taught me a lot while growing up together. Thank you for all the many different experiences and joyful moments. You were always a great source of inspiration. I also want to thank my late grandparents, Anni & Tibor and Emma & Ludwig, for teaching me curiosity, kindness, and generosity. I fondly remember their warmth, interest, and compassion whenever we spent time together. Especially, granddad Tibor's stimulating questions and riddles nurtured my curiosity already at a very young age. Why does a pond freeze from above and not below? How many times do glasses clink if a group of N people come together? If there is one road to freedom and another to damnation, and two guards – one of them always tells the truth and one of them always lies – and I do not know who is who: how should one single question be formulated to know with certainty which road to take? These types of questions were always challenging but showed me what kind of satisfaction one can get from discovering right answers. Moreover, I am grateful to all other members of my big family, who were a constant source of joy and comfort.

Finally, I want to thank my wife Sasha for the unconditional love and for seeing and nurturing me for who I am. I would not be the person that I am today without you, and this dissertation would certainly look very different without your input. Thank you for my immense personal growth, for believing in me – even in the most difficult times – for teaching me the ability to stand up for myself and the values that are important for me. For seeing the world how it is, and pushing for how it should be. For always being there for me, for countless inspiring discussions, our deep conversations, for the silliness, and for your trueness and kindness. I am and always will be grateful for you – without measure.

Contents

List of Figures

List of Tables

Introduction

This dissertation consists of three independent chapters, which are broadly connected to destructive behavior in competitive settings and ways how to mitigate it. Specifically, in Chapter 1, I analyze contests, where players can sabotage each other and investigate whether a designer should disclose the number of competitors when there is uncertainty about the group size. Chapter 2 focuses on a competitive litigation setting. It investigates whether and how spiteful preferences contribute to excessive litigation expenditures and how the choice of the fee-shifting rule can mitigate such behavior. Chapter 3 moves from destructive behavior in competitive settings to analyzing a mitigation mechanism for selfish behavior more broadly. In particular, it explores the occurrence of informal punishment from third parties and studies whether perceptions of social norms are motives for such punitive actions. While it does not study this mechanism specifically for destructive behavior in competitive settings, it nonetheless can inform the occurrence of informal sanctioning as a mitigation mechanism in such environments.

In Chapter 1, I focus on contests, where competitors can not only engage in constructive effort but in sabotage as well. In many contests, players are not aware of how many competitors they face. While existing studies examine how disclosing this number affects participants' productive effort, this paper is the first to consider its impact on destructive behavior. To do so, I theoretically and experimentally study how revealing the number of contestants affects both effort and sabotage compared to concealing this information. Further, I evaluate the created value by comparing the resulting performances, which are shaped by the combination of the exerted effort and the received sabotage. The results show that the overall performance can be higher under concealment, even though the disclosure policy does not affect average effort and sabotage levels. That is because the distribution of effort and sabotage differs between the disclosure policies: when players know how many competitors they are facing, they can adjust their effort and sabotage levels to that specific group size, whereas they have to choose one effort and one sabotage for all group sizes when they do not have this information. The experimental results largely confirm these theoretical predictions and demonstrate the significance of accounting for the effects of sabotage, as it induces performance differences between the group size disclosure policies. By concealing the number of contestants, a designer can mitigate the welfare-destroying effects of sabotage, without curbing the provision of value-creating effort.

In Chapter 2, together with Wladislaw Mill, we focus on a litigation setting, where the plaintiff competes with the defendant to win a legal case. It is known that some litigants engage in overly excessive litigation expenditures, which, from a society's point of view, can be considered wasted resources. We explore whether spiteful preferences motivate increased litigation spendings and how the choice of the fee-shifting rule can mitigate such behavior. Under the American fee-shifting rule, both players have to pay for their own legal costs, independent of who wins the case. Under the English rule, the loser also has to the winner's expenditures. Additionally, we investigate the effect of spiteful preferences and the fee-shifting rule on pre-trial settlement behavior. To do so, we derive theoretical predictions and test them with an online experiment. We find that litigation expenditures are overall higher under the English rule compared to the American – even for low-merit cases – while there is no difference for settlement requests. Spiteful participants exhibit overall higher expenditures and settlement requests, with a more pronounced increase in litigation expenditures under the American fee-shifting rule. The increase in settlement requests is similar under both rules. Our results indicate that being spiteful does not pay off in monetary terms. The expected payoff is lower for more spiteful litigants – especially under the American rule – independent of facing a less or more spiteful opponent. Moreover, being matched with a more spiteful litigant reduces the expected payoff similarly under both rules. We conclude that the English rule can protect spiteful players from lowering their own expected payoffs but cannot reduce the harm they inflict upon others – at the cost of inducing higher litigation expenditures for less spiteful players compared to the American rule.

In Chapter 3, together with Katarína Čellárová, we focus on costly punishment from an unaffected third party, which can play an important role not only in sustaining cooperation but also in deterring selfish and destructive behavior. In this chapter, we want to understand better its occurrence and study perceptions about social norms as the underlying motives. In the literature, such third-party punishment has been taken as evidence in itself that individuals care about the enforcement of social norms. In Chapter 3, we explicitly study whether and which norm-related beliefs motivate third-party punishment. To do so, we run an experiment where we elicit punishment decisions in a modified dictator game and measure three social norm perceptions: personal norms of appropriateness, beliefs about others' appropriateness norms (normative expectations), and beliefs about typical behavior (empirical expectations). We find that higher personal norms of appropriateness and higher empirical expectations lead to an increase in punishment. Normative expectations, on the other hand, are negatively correlated with punishment when controlling for either of the other two norm perceptions. We conclude that the desire to enforce own beliefs of appropriateness or typical behavior motivates punishment decisions rather than perceived societal appropriateness views.

Chapter 1

Disclosure Policy in Contests with Sabotage and Group Size Uncertainty

1.1 Introduction

Contests exist in many settings, including job promotion tournaments, crowdsourcing contests, academic research grant applications, and procurement auctions. In these competitive situations, agents spend non-refundable resources to outperform one or more competitors to enhance their chances of winning a valuable prize. However, in many cases, agents are not aware of how many other contestants they are facing, and whether there is another competitor at all (e.g., Boosey et al., 2017; Morgan et al., 2012; Lim and Matros, 2009). In those cases, a contest designer, seeking to increase value-creating effort provisions, may decide to disclose the number of contestants or leave uncertainty about the group size. For instance, in a workplace setting, a manager may decide to reveal the number of short-listed candidates being considered for promotion. Similarly, when companies or the government offer inducement prizes for innovations or conduct procurement auctions, they may choose to disclose the number of participating competitors.[1]

For contest designers, disclosing the number of participants is an easy-to-implement tool. For contestants, this decision can have implications for their effort levels. That is because winning chances are determined by contestants' performances *relative* to the performances of their competitors. Relative performances are shaped by each contestant's effort level and, thus, deciding how much effort to exert depends on the number of competitors and beliefs about their effort levels. In line with theoretical equilibrium predictions, the experimental literature shows that if contestants know how many other contestants there are, effort usually decreases with an increasing group size (Dechenaux et al., 2015). If they do not know

[1] See (Fu et al., 2016) for a more detailed discussion of these examples.

this number, it becomes more difficult to determine how much effort is needed to outperform others. In such cases, the theoretical equilibrium decisions are a weighted sum of the equilibrium choices conditional on the group sizes (Lim and Matros, 2009), which lead to no difference in the average effort levels between disclosing and concealing the number of contestants (Fu et al., 2016, 2011). In these standard settings, even though effort choices become more difficult when players do not know the number of competitors, the experimental literature is consistent with theory as it does not find significant differences between the two disclosure policies (Jiao et al., 2022; Boosey et al., 2020; Aycinena and Rentschler, 2019).

Yet, the choice of the disclosure policy may not only influence contestants' constructive efforts, but also induce destructive behavior such as sabotage. Along with effort, sabotage is another strategy to increase one's relative performance – not through own productive effort but by negatively distorting one's competitors' performances, and a substantial literature has emerged on this topic (e.g., Chowdhury et al., 2023; Dato and Nieken, 2020; Chowdhury and Gürtler, 2015; Charness et al., 2014; Gürtler et al., 2013; Harbring and Irlenbusch, 2011; Carpenter et al., 2010; Lazear, 1989). Such sabotage can take various different forms. For instance, in workplace promotion tournaments, co-workers may withhold important information, skills, or experiences, share only partial information, or even provide wrong information to reduce the productivity of their colleagues (e.g., Serenko, 2020; Pan et al., 2018; Kumar Jha and Varkkey, 2018; Evans et al., 2015; Ford and Staples, 2010). Sabotage can also occur between companies, for instance through cyberattacks on the information and production systems of potential competitors. Bitkom (2018) estimated that in Germany alone in the years 2017 and 2018, more than four billion Euros were destroyed because of cyberattacks as a form of sabotage between competing companies. If such destructive behavior happens, effort is spent less productively, which results in a decrease in the overall created value. For instance, sabotaged co-workers may work with less efficient tools and focus on less important tasks, or companies have to spend their resources to fix the created damages.[2]

Although sabotage is of such importance for welfare, we still do not know how the choice of a group-size disclosure policy affects sabotage behavior. Yet, a policy that aims to increase welfare should take the adverse effects of sabotage into account. In this paper, I address this research gap by theoretically modeling and experimentally testing the differences between concealing and disclosing the number of contestants, taking into account not only contestants' effort choices but their sabotage decisions, as well. I first analyze the comparative statics of the realized group sizes under disclosure and the comparative statics of different enter

[2]Other common sabotage examples include the denigration of potential competitors' products or services (Nissen and Haugsted, 2020), negative campaigning in political races (Lau and Rovner, 2009), or fouls in sports (Deutscher et al., 2013). In this paper, I focus on sabotage that is used to decrease the productivity of competitors and thus destroys value.

probabilities and number of potential contestants under concealment. Then, I compare the resulting efforts and sabotage levels, expected payoffs, and performances between the disclosure policies. As a welfare measure, I focus on the sum of individual performances (*group performance*), as it shows the overall created value in the presence of sabotage-induced value losses.

In my theoretical analysis, I follow Konrad (2000) to model sabotage in a Tullock contest (Tullock, 1980) and employ exogenous enter probabilities to model group size uncertainty, following Lim and Matros (2009). I introduce a designer, who commits to always conceal or disclose the number of contestants, but not their identities, following Fu et al. (2011). The number of potential contestants (and their enter probabilities) are common knowledge. As a consequence, players can sabotage all those that potentially compete with them independent of whether they know the number of actual competitors or whether there is uncertainty about it. For example, in a workplace context, players may have a sense of who potentially also applies for a position based on the position's requirements, allowing them to (preemptively) sabotage all of them. This sabotage could include not sharing crucial information or skills, or even providing wrong information and advice. Thus the knowledge of the set of potential competitors allows contestants to sabotage all of them, irrespective of the disclosure policy, as the designer merely discloses the number of active contestants, but not their identities. Similar dynamics can arise in procurement auctions or other contests, where there are well-defined sets of 'usual suspects'.

The theoretical results show that average effort and sabotage levels, as well as expected payoffs, are not different between the disclosure policies. However, the average group performance is higher under group size concealment compared to a disclosure of the realized group sizes. This is because group performance increases in own effort levels and decreases in the received sabotage. Contestants can adjust their effort and sabotage level to the specific realized group size under disclosure, while they have to choose one effort and one sabotage level, which will be used for any realized group size when it is concealed. Consequently, the distribution of effort and sabotage across group sizes differs, which induces differences in the group performances. The highest performance difference occurs in the case when there is only one contestant, who will win the prize with certainty. In this case, the one contestant does not exert any effort, when knowing to be the only contestant, compared to taking into account also other possible group size realizations, when not knowing the realized group size. This leads to the exertion of a substantive amount of effort in anticipation of other group size realizations and receiving sabotage, while actually being the only contestant and not receiving any sabotage. Hence, the resulting performance is particularly high, which shapes the overall increase in group performance under group size concealment compared to disclosure. This is because for all other group size realizations, the performance differences between the disclosure policies are small, resulting in a higher average performance under concealment when there is at least a 1 percent chance of being the only contestant. To evaluate the validity of these theoretical predictions, I conduct an experiment[3]. As sabotage is difficult to observe in the field,[4] a laboratory experiment is an optimal environment to test theories involving the possibility of sabotage. This holds especially true for this paper's setting because it involves a complex setting with several sources of uncertainties and best responses to competitors' effort *and* sabotage levels.[5]

In the experiment, subjects play a Tullock contest with group size uncertainty, where each group member has the same exogenous enter probability. I vary the disclosure policy (concealment vs. disclosure) within subjects, and enter probabilities (0.25 vs. 0.75) and the size of the group (3 vs. 5) between subjects, leading to different probabilities of being the only contestant (0.4%, 6%, 32%, and 56%). Subjects receive an endowment that they can use to invest in 'Option A' (effort) to improve their own performance, or in 'Option B' (sabotage) to negatively affect everyone else's. Under group size disclosure, subjects make effort and sabotage decisions conditional on the realized group size via the strategy method, whereas under concealment they make one effort and one sabotage decision to fit all realized group sizes. To create the notion of value-creating effort and value-destroying sabotage in the experiment, money is donated to a non-profit charity, and the amount depends on the absolute performance of the group. Hence, by investing in effort, subjects increase both their own performance and donations, whereas by investing in sabotage, they increase their own relative performance by decreasing their opponents' performances but at the cost of decreasing donations. Importantly, the inclusion of this externality does not change the theoretical predictions, even if subjects have a preference for donations.

The experimental results are largely in line with theory and add to our understanding of contestants' behavior under the two disclosure policies. The first key finding is that group performance is significantly higher under concealment compared to disclosure but only when the probability of being the only contestant is not too low. As predicted, this difference is driven by the possibility of being the only contestant, where subjects do not

[3]The experiment was preregistered at aspredicted.org `https://aspredicted.org/blind.php?x=VB2_4DF` and received ethical approval from the Ethics Committee of the University of Mannheim.

[4]Observational studies usually rely on sports data to identify sabotage, which they typically define as the breaking of rules (e.g. Brown and Chowdhury, 2017; Deutscher et al., 2013; Balafoutas et al., 2012; Del Corral et al., 2010).

[5]As a consequence, behavior may be influenced by other factors such as bounded rationality, probability distortions, risk aversion, and many others. Additionally, contests typically also induce non-monetary utilities such as joy of winning, which can lead to heterogeneous behavior (Dechenaux et al., 2015). With the existence of sabotage, other motives such as spitefulness may become relevant. Therefore, this experiment can be viewed as a robustness test for the theoretical predictions, which allows for these additional factors.

receive any sabotage and exert much higher effort under concealment compared to disclosure. Consequently, there is no difference in group performance, when the probability of being alone is 0.4%, but in all other treatments where this probability is at least 6%, concealment leads to a higher group performance.

The second key finding is that there is no evidence for a difference in average sabotage and effort levels, as well as in expected payoffs between the two disclosure policies. The only exception is when the number of potential contestants is 3 and enter probabilities 0.25. In this case, concealment leads to a slight increase in sabotage levels. Nonetheless, even in this case, the expected payoffs do not differ between the disclosure policies.

As additional results, I confirm the predicted comparative statics of disclosed group sizes, where a larger group size reduces sabotage and effort levels. At the same time, there is above-equilibrium sabotage in groups of sizes 3, 4, and 5. This behavior can be explained by joy of winning that increases in the number of competitors (constant winning aspiration) (Boosey et al., 2017), or by spiteful preferences (Morgan et al., 2003a; Levine, 1998). As to the comparative statics of group size uncertainty, I find that an increase in the number of potential contestants decreases sabotage levels for high enter probabilities, as theory suggests. For low enter probabilities, however, I do not find evidence for the hypothesized increase.

The contribution of this paper is as follows. I add to the discussion of group size disclosure policies, by examining contestants' behavior in a more nuanced setting, that allows not only for constructive behavior but also for destructive behavior. The literature shows that competition also induces cheating, fraud, and sabotage besides productive efforts (Piest and Schreck, 2021; Chowdhury and Gürtler, 2015; Carpenter et al., 2010; Faravelli et al., 2015), and thus a more realistic contest setting should account for such behavior. Moreover, the inclusion of sabotage is indispensable for policy evaluations, as sabotage destroys value and therefore has negative welfare implications. In the most standard contest setting without sabotage, the disclosure policy does not influence the average exerted effort and hence the created value (Lim and Matros, 2009). I show that when sabotage in contests is accounted for, higher performances can be induced by concealing the number of competitors. This has substantial implications for contests' design. A designer can mitigate the welfare-destroying effects of sabotage by concealing the number of contestants.

I also contribute to the sabotage literature by suggesting a policy that mitigates the destructive effects of sabotage without curbing productive efforts. The theoretical and experimental literature shows ways of how to decrease sabotage altogether, including reducing the prize spread (Harbring and Irlenbusch, 2011, 2005; Del Corral et al., 2010; Vandegrift and Yavas, 2010; Lazear, 1989), increasing the number of contestants (Konrad, 2000), increasing the penalties for sabotage (Balafoutas et al., 2012), revealing the identity of the saboteur

(Harbring et al., 2007), or not revealing intermediate relative performances or rank (Charness et al., 2014; Gürtler et al., 2013; Gürtler and Münster, 2010) as sabotage is directed against the most able or best-performing contestant (Deutscher et al., 2013; Harbring et al., 2007; Münster, 2007; Kräkel, 2005; Chen, 2003). For broader literature reviews on sabotage in contests see Piest and Schreck (2021); Amegashie et al. (2015), or Chowdhury and Gürtler (2015).

This paper also informs other theoretical contest settings without sabotage, where there are already differences in effort choices between the two group size disclosure policies. Accounting for the effects of sabotage may interact with their identified effects and possibly change the conclusions. These settings include different prize valuations together with different enter probabilities (Fu et al., 2016), different prize valuations with endogenous entry (Chen et al., 2023), the existence of bid caps (Wang and Liu, 2023; Chen et al., 2020a), either convex or concave cost structures (Jiao et al., 2022; Chen et al., 2017), and either strictly convex or concave characteristic functions of the Tullock contest (Feng and Lu, 2016; Fu et al., 2011).

I further add to the experimental contest literature without sabotage (Jiao et al., 2022; Boosey et al., 2020; Aycinena and Rentschler, 2019), which, in most settings, finds no difference in average effort levels between the two disclosure policies. In more specific settings, the experimental literature finds that disclosure can lead to higher effort levels, for instance when the outside option is high and entry endogenous (Boosey et al., 2020), or when effort costs are concave (Jiao et al., 2022). In this paper, I show that concealment leads to a higher performance, even though there are no differences in the average effort and sabotage levels.

By also studying the comparative statics of group size, I provide evidence for the influence of known group sizes on sabotage, which so far lacks empirical evidence as pointed out by Piest and Schreck (2021) and Chowdhury and Gürtler (2015).[6] As I find substantial over-sabotage for larger group sizes, I argue that sabotage is not necessarily a 'small number phenomenon' (Konrad, 2000). Therefore, increasing group size may not be an apt tool to decrease overall sabotage and should therefore be used with caution, if at all.

Moreover, my paper is the first to consider group size uncertainty in a contest with sabotage. For contests without sabotage, the literature shows that group size uncertainty matters for effort levels of contestants (Gu et al., 2019; Boosey et al., 2017; Chen et al., 2017; Ryvkin and Drugov, 2020; Kahana and Klunover, 2016, 2015; Morgan et al., 2012; Fu et al., 2011; Lim and Matros, 2009; Münster, 2006; Myerson and Wärneryd, 2006; Higgins et al., 1988). Yet, the existing sabotage literature assumes that the number of contestants

[6]So far, there is only one experimental study that investigates a known number of competitors but in a rank-order tournament, which predicts no differences in sabotage levels across group sizes. Thus, the authors do not find any differences in their experiment (Harbring and Irlenbusch, 2008).

is common knowledge.[7] I experimentally confirm that effort and sabotage decisions under uncertainty can be described by a weighted sum of the level choices for the known group sizes.

The structure of this paper is as follows: In Section 3.2, I set up a theoretical model in order to derive equilibrium predictions. Section 1.3 describes the experimental design. In Section 3.4, I present the results before I provide a discussion and conclusion in Section 1.5.

1.2 Theoretical Model and Predictions

In this section, I introduce the theoretical model, which guides the experimental analysis. I also shortly introduce the experimental setting and derive hypotheses.[8]

1.2.1 Setup

I follow Konrad (2000) to model sabotage in a Tullock contest (Tullock, 1980) and employ exogenous enter probabilities to model group size uncertainty, following Lim and Matros (2009).[9]

Let N be the set of all homogenous and risk-neutral potential contestants, and n the number of potential contestants indexed by $i \in N, N = \{1, ..., n\}$. Every potential contestant has the same enter probability of $q \in (0, 1]$. The set of potential contestants N and their enter probabilities q are common knowledge. Let N_i be the set of possible opponents of player i. Conditional on player i participating, let $M_i \subseteq N_i$ be the set of other active players except for player i in the contest. M_i is not known to the players. Let m be the number of active contestants including player i with $M = \{1, ..., m\}$ being the set of all active contestants including player i.

There is a contest designer, who ex-ante commits to always conceal or reveal the number of active contestants m.[10] She does not reveal the identities of the active players. Because

[7] Chowdhury et al. (2023, 2022); Dato and Nieken (2020, 2014); Benistant and Villeval (2019); Brown and Chowdhury (2017); Leibbrandt et al. (2017); Charness et al. (2014); Deutscher et al. (2013); Gürtler et al. (2013); Amegashie (2012); Balafoutas et al. (2012); Harbring and Irlenbusch (2011); Carpenter et al. (2010); Vandegrift and Yavas (2010); Gürtler and Münster (2010); Harbring and Irlenbusch (2008); Münster (2007); Harbring et al. (2007); Kräkel (2005); Chen (2003); Konrad (2000); Lazear (1989).

[8] The hypotheses are pre-registered on `https://aspredicted.org/VB2_4DF`.

[9] Exogenous enter probabilities may arise when a contest is exposed to specific regulations and entry barriers, that include certain quality and safety standards of a product in a patent race, specific requirements concerning skills and characteristics of employees for a promotion, or legislation designing lobbying rules (Boosey et al., 2017). Likewise, they may arise as mixed-strategy equilibrium enter choices (Fu et al., 2015) determined by the value of the prize, entry fees, and the outside option.

[10] If the designer decides to partially disclose the number of contestants, contestants can anticipate the specific realized group sizes, where the designer would prefer to disclose. Lim and Matros (2009) show in a contest without sabotage, that if a designer can not credibly commit to always either conceal or disclose, she would always disclose the number of contestants. Similar dynamics would arise in this more specific setting

of this, players can only choose to sabotage all other potential contestants N_i, because they know who potentially enters, but they do not know who actually entered. I assume that they can only sabotage all others the same amount.[11]

Active players compete to win a single prize W. They choose to spend effort $e_i \geq 0$ with linear costs $C(e_i) = e_i$ and sabotage $s_i \geq 0$ with linear sabotage costs $C(s_i) = s_i$.[12] Contestant i is subjected to total sabotage of $\sum_{j \in M_i} s_j$. Only active players are affected by the exerted sabotage as only they exert contest-induced additional efforts.[13] The effort and sabotage levels translate into individual performance y_i as follows:

$$y_i = \frac{e_i}{1 + \sum_{j \in M_i} s_j}$$

Individual performances are increasing in contestants' own effort levels and decreasing in the total amount of received sabotage (i.e., their opponents' sabotage levels).[14] Player i's probability of winning is determined by the following contest success function:[15]

$$p_i(y_i, y_{-i}, M_i) := \begin{cases} \frac{y_i}{y_i + \sum_{j \in M_i} y_j} & \text{if } \max\{y_1, ..., y_m\} > 0 \\ \frac{1}{m} & \text{otherwise,} \end{cases}$$

With this contest success function, relative performances determine individual winning probabilities. Therefore, players have two options to increase their winning chances. They can either increase their own performance by providing additional effort or decrease their opponents' performances by sabotaging more. An essential feature of group size uncertainty is the possibility of being the only contestant. In this case, the one only active player i wins the contest with certainty independent of her effort and sabotage choices.

but is beyond the scope of this paper.

[11] Sabotaging all others the same amount would arise in equilibrium when contestants are homogenous and could decide to individually sabotage others. As players do not know the identities of the active contestants, even under disclosure, there is no benefit in sabotaging only one other player, because it would introduce a coordination problem with the other players.

[12] Sabotage costs incorporate expected punishment costs and reputation losses for detected sabotage, possible moral costs, costs for hiding the exerted sabotage, and possible long-run costs, for example, when sabotage decreases the future productivity of agents.

[13] This assumption isolates the effect of the disclosure policy on the contest-induced performances. Additionally, if the sabotage is specifically directed towards only contest-related efforts, such as withholding information about promotion-relevant work activities, there is no effect on non-active players. Even if there is an effect on the base productivity of non-active players, and this base productivity is small enough or the effectiveness of sabotage on this base productivity is small, the results remain the same. See Section 1.5 for a more detailed discussion.

[14] The results extend to performance functions with less pronounced marginal returns in the received sabotage: $y_i = \frac{e_i}{(1+\sum_{j \in M_i} s_j)^t}$ with $t < 1$. For $\lim_{t \to 0}$, however, sabotage has no effect anymore and the performance differences between the disclosure policies disappear. See Appendix A.1.6 for a more detailed analysis.

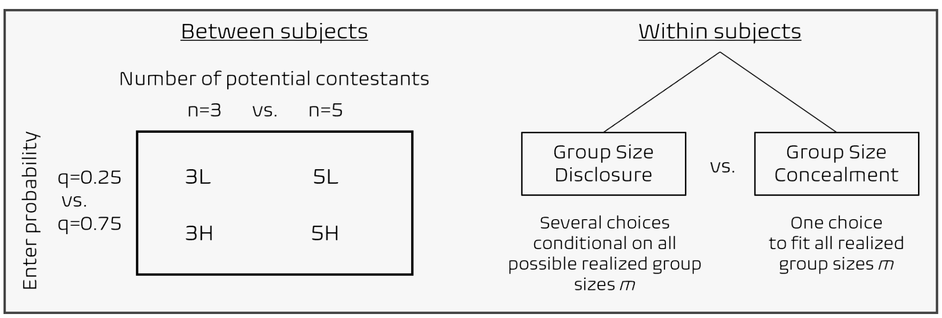
[15] For an axiomatization see Skaperdas (1996).

The timing of the game is as follows. Before the contest, the designer ex-ante commits to always conceal or disclose the number of contestants. Then, nature determines who becomes active and enters the contest. Conditional on participating, active contestants simultaneously make their effort and sabotage choices. Afterwards, the contest is resolved according to the winning probabilities.

1.2.2 Experimental Conditions

Figure 1.1 shows an overview of the experimental conditions. I exogenously vary the number of potential contestants from $n = 3$ to $n = 5$ and enter probabilities from $q = 0.25$ to $q = 0.75$ between subjects, resulting in the treatments $3L$, $5L$, $3H$, and $5H$ with different probabilities of being the only contestant (56% in *3L*, 32% in *5L*, 6% in *3H*, and 0.4% in *5H*). At the same time, I vary the disclosure policy within subjects, hence, every subject makes decisions both under group size disclosure and group size concealment. See Section 1.3 for the full description of the experiment.

Figure 1.1: Experimental conditions



1.2.3 Group Size Disclosure

Under group size disclosure, the designer ex-ante commits to reveal the number of contestants, however not their identities. Therefore, players do not know who exactly are their competitors, but they know the number of active contestants and the set of all other potential contestants N_i. As a consequence, they can sabotage all other potential contestants, which include their actual competitors. The decision how much effort and sabotage to exert is therefore based on their strategic response to the number of competitors and their effort and sabotage levels. Conditional on being active, player i chooses e_i and s_i to maximize her expected payoff:

$$\arg\max_{e_i, s_i} p_i(y_i, y_{-i}, m)W - e_i - s_i \tag{1.1}$$

The associated first-order and second-order conditions can be found in Appendix A.1.1. Conditional on being active, all contestants simultaneously choose effort and sabotage. The following proposition characterizes the static symmetric equilibrium:

Proposition 1. *Consider a contest as described above. The static symmetric equilibrium is characterized as follows:*

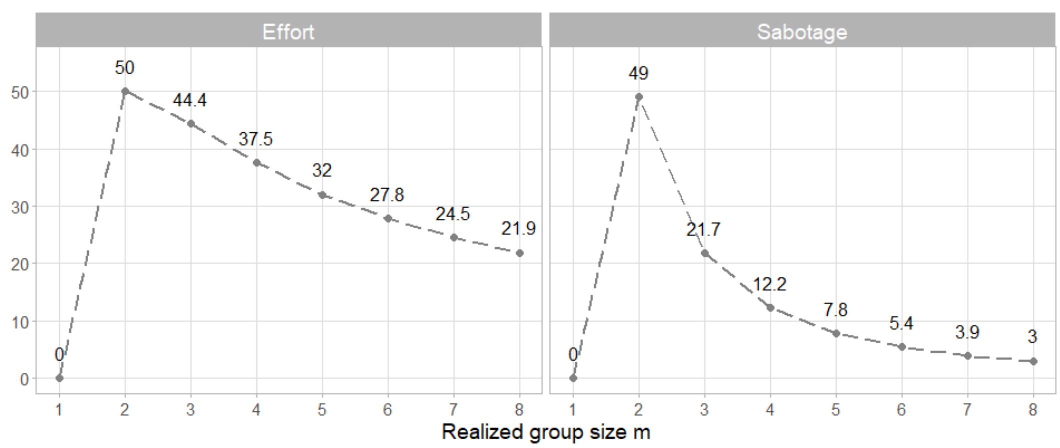
$$e^* = \frac{(m-1)}{m^2}W \tag{1.2}$$

$$s^* = \begin{cases} \frac{1}{m^2}W - \frac{1}{m-1} & \text{if } W \geq \frac{m^2}{(m-1)} \text{ and } m \geq 2 \\ 0 & \text{else} \end{cases} \tag{1.3}$$

Proof. See Appendix A.1.2 □

Figure 1.2 shows the static symmetric equilibrium for effort and sabotage levels depending on the realized group size m. It includes the case when there is no other competitor ($m = 1$). In this case, the one active player wins the prize with certainty, making it optimal to not exert any effort or sabotage. When there is at least one other contestant ($m > 1$), equilibrium effort and sabotage levels decrease with increasing group size due to more competition. Sabotage is impacted more than effort due to the additional *dispersion effect* (Konrad, 2000). Any sabotage against one player benefits all other players, and hence players can free-ride on their competitors' sabotage levels. With more opponents, these dispersion effects increase, and their own exerted sabotage becomes relatively less beneficial.[16] Following the theoretical model, I hypothesize the following:

Figure 1.2: Static symmetric equilibrium effort and sabotage under group size disclosure



Note: The figure depicts equilibrium effort and sabotage levels conditional on the realized group size m for a prize of $W = 200$

[16]The dispersion gains also exist when all competitors are sabotaged simultaneously, as one agent still profits from the sabotage against the others.

Hypothesis 1.1. *A larger disclosed group size decreases effort and sabotage levels for $m > 1$.*

1.2.4 Group Size Uncertainty

Under group size uncertainty, the contest designer ex-ante commits to conceal the number of contestants. Hence, contestants do not know the number of active contestants. Instead, they know the set of all other potential contestants N_i and their enter probabilities q. With these, they can compute the expected number of contestants. Because they know the identities of every potential contestant, as under group size disclosure, active players can exert sabotage against all other potential contestants. Hence, conditional on being active, player i chooses e_i and s_i as follows:

$$\arg\max_{e_i,s_i} \sum_{M_i \in \mathcal{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|} p_i(y_i, y_{-i}, M_i)W - e_i - s_i \tag{1.4}$$

where $\mathcal{P}^{N_i}$ is the powerset of N_i. Conditional on participating, players simultaneously maximize their expected profit function by choosing e_i and s_i. The following proposition characterizes the static symmetric equilibrium:

Proposition 2. *Consider a contest with group size uncertainty as described above. Conditional on being active, the optimal effort in the static symmetric equilibrium is described by:*

$$e^* = \sum_{(m-1)=0}^{n-1} \underbrace{\frac{(n-1)!}{(m-1)!(n-m)!} q^{m-1}(1-q)^{n-m}}_{\textit{probability of } m-1 \textit{ others}} \times \underbrace{\frac{m-1}{m^2}W}_{\textit{effort choice for } m-1 \textit{ others}} \tag{1.5}$$

A numerical solution to the following equation describes the optimal sabotage level s^*:

$$\sum_{(m-1)=0}^{n-1} \underbrace{\frac{(n-1)!}{(m-1)!(n-m)!} q^{m-1}(1-q)^{n-m}}_{\textit{probability of } m-1 \textit{ others}} \times \frac{m-1}{m^2} \frac{1}{1+(m-1)s} W = 1 \tag{1.6}$$

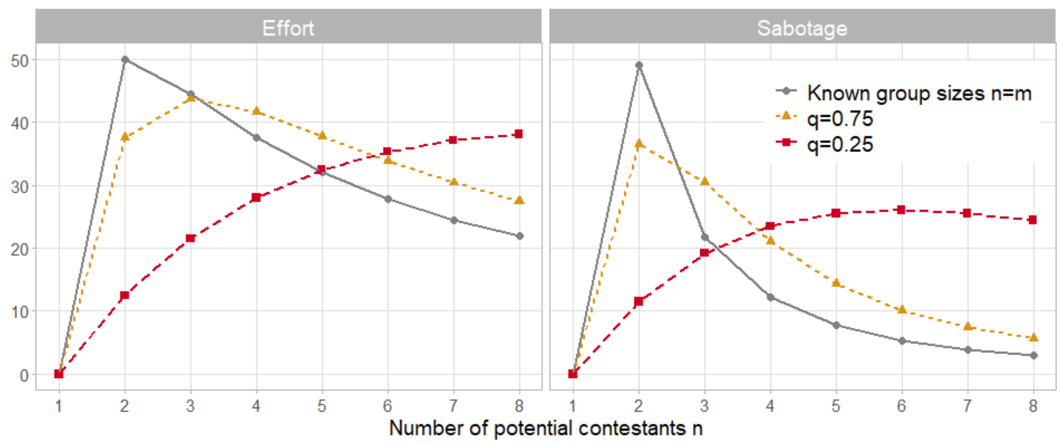
Proof. See Appendix A.1.3 □

Proposition 2 shows that effort decisions under group size uncertainty are a weighted sum of the equilibrium choices for known realized group sizes. For sabotage, there is a numerical solution, but the choices are almost the weighted sum of the equilibrium choices for known realized group sizes.[17] Figure 1.3 depicts the comparative statics of group size uncertainty and shows the influence of the number of potential contestants n and their enter probabilities

[17] There is no closed form solution because in the performance function, 1 is added to the received sabotage to ensure a solution in the special case of not receiving any sabotage $y_i = \frac{e_i}{1+\sum_{j \neq i} s_j}$.

($q = 0.25$ vs. $q = 0.75$) on equilibrium effort and sabotage levels. Additionally, it depicts the equilibrium choices for known group sizes to illustrate that effort and sabotage choices under group size uncertainty are the weighted sum of the equilibrium choices under disclosure. As a consequence, an interesting change in the comparative statics of the potential number of contestants n arises. Specifically, when enter probabilities are high ($q = 0.75$), sabotage decreases when the number of potential contestants n increases from 3 to 5, whereas when enter probabilities are low ($q = 0.25$), sabotage increases. Hence, for the specific conditions in the experiment, I hypothesize the following:

Figure 1.3: Static symmetric equilibrium effort and sabotage under group size concealment



Note: The figure depicts equilibrium effort and sabotage levels for uncertain group sizes for enter probabilities of 0.75 (yellow) and 0.25 (red). Additionally, it depicts the comparative statics of known group sizes (where the y-axis becomes the realized group size m). The prize is set to $W = 200$.

Hypothesis 1.2. *For high enter probabilities ($q = 0.75$), effort and sabotage levels decrease when the number of potential contestants increases from $n = 3$ to $n = 5$.*

Hypothesis 1.3. *For low enter probabilities ($q = 0.25$), effort and sabotage levels increase when the number of potential contestants increases from $n = 3$ to $n = 5$.*

1.2.5 Comparing Disclosure Policies

In the following, I compare the effects of the disclosure policy on expected effort and sabotage levels, as well as on expected payoffs. Additionally to expected payoffs, I consider the expected sum of individual performances as a welfare measure, because it incorporates the value-creating effects of effort and the value-destroying effects of sabotage.

Expected Effort, Sabotage, and Payoffs

When there is uncertainty about the number of active contestants, contestants take the weighted sum of their equilibrium effort levels for the known group sizes. The expected effort is the same value, as there is only one effort choice for all realized group sizes. Under disclosure, the expected effort is the exact same weighted sum. As a consequence, there is no difference in the expected effort between disclosing and concealing the number of contestants (see Appendix A.1.4). Moreover, a numerical analysis shows that there are also no substantial differences in sabotage levels (see Appendix A.1.4).

Hypothesis 2.1. *There are no substantial differences in expected effort and expected sabotage levels between concealing and disclosing the number of contestants.*[18]

The expected costs are the same across the disclosure policies because there is no difference in the expected effort and sabotage levels. Additionally, in the symmetry equilibrium, everyone exerts the same amount of effort and sabotage, leading to the same winning probabilities independent of the realized group size and policy. Consequently, there is no difference in the expected payoffs between the disclosure policies (see Appendix A.1.5).

Hypothesis 2.2. *There is no substantial difference in expected payoffs between disclosure and concealment.*

Expected Group Performance

Next, to compare the created value, I study the differences in the expected sum of individual performances (group performance) between the disclosure policies. For this, I first study group performance conditional on the realized number of contestants m. Under group size disclosure, players can adjust their effort and sabotage levels according to the realized group size m ($e^*(m)$, $s^*(m)$). Under group size concealment, contestants cannot do this and have to choose one effort and one sabotage level for all realized group sizes ($e^*(n,q)$, $s^*(n,q)$). The equilibrium group performance conditional on the realized group size m can be described as follows:

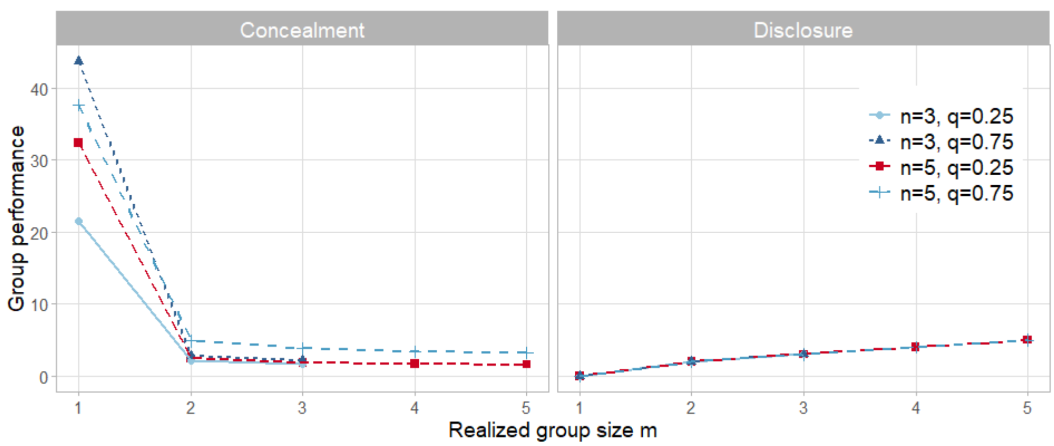
$$P_{disclosure}(m)^* = \underbrace{\sum_{i=1}^{m} y_i(m)}_{\text{sum of individual performances}} = \underbrace{\frac{e^*(m)}{1+(m-1)s^*(m)}}_{\text{individual performance}} \times \underbrace{m}_{\text{realized number of contestants}} \tag{1.7}$$

[18]This hypothesis was not preregistered and was added later. However, it follows directly from the model that remained unchanged.

$$P_{concealment}(m)^* = \underbrace{\sum_{i=1}^{m} y_i(m,n,q)}_{\text{sum of individual performances}} = \underbrace{\frac{e^*(q,n)}{1+(m-1)s^*(q,n)}}_{\text{individual performance}} \times \underbrace{m}_{\text{realized number of contestants}} \tag{1.8}$$

Figure 1.4 depicts these equilibrium group performances, conditional on the realized number of contestants m, and the treatments (combinations of the number of potential contestants n and their enter probabilities q). When the number of contestants is disclosed (right panel), each individual's equilibrium performance is exactly 1 for $m > 1$. As the number of contestants m increases, the group performance increases because the individual performances are summed up. When the contestant is alone in the contest ($m = 1$), she does not exert any effort, resulting in a performance of 0.

Figure 1.4: Equilibrium group performance per disclosure policy and realized group size



Note: The figure illustrates the equilibrium group performance (sum of individual performances) conditional on the realized number of contestants under concealment (left graph) and disclosure (right graph). The different colors indicate the between treatments. Under disclosure, all four lines are exactly the same. The prize is $W = 200$.

When the number of contestants is concealed (left panel), contestants cannot adjust their effort and sabotage levels to the realized group size. Instead, they choose one effort and sabotage level that is used for all realized group sizes. As a consequence, larger groups suffer from more sabotage overall, while the amount of effort stays constant. Therefore, individual performances and even group performances fall in the group size. A special case is $m = 1$ when a player is the only contestant. In this case, she does not receive any sabotage while exerting a substantive amount of effort, leading to a particularly high performance also because of decreasing marginal returns of the received sabotage.[19] This performance

[19]This difference is also pronounced for performance functions that have a less pronounced decrease in

is substantially higher than the performance for any other realized group size and all other performances under group size disclosure.

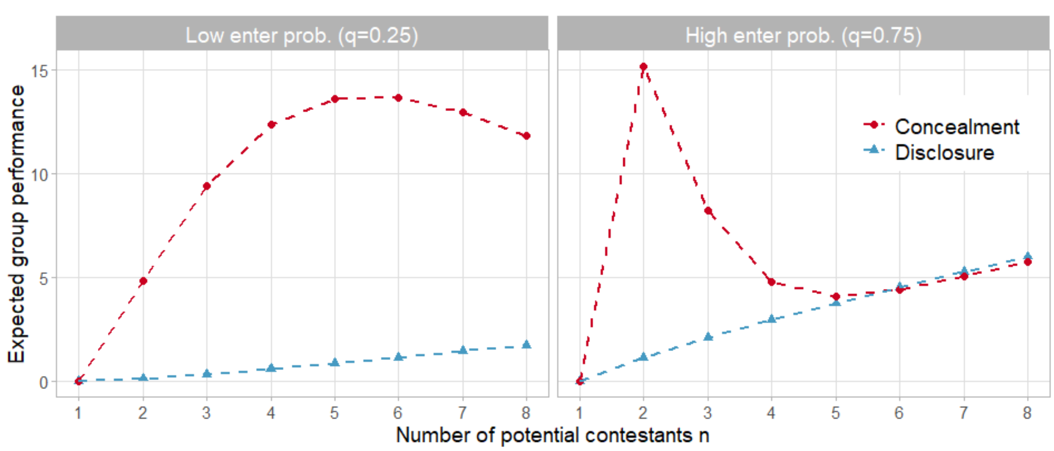
Next, I compare the resulting expected total group performance conditional on the number of potential contestants n and their enter probability q. The expected group performance is a weighted sum over all group size realizations and their specific group performance:

$$E[P_{Disclosure}(m)] = \sum_{m=1}^{n} \underbrace{\frac{n!}{m!(n-m)!} q^m (1-q)^{n-m}}_{\text{probability of group size } m} \times \underbrace{\frac{e^*(m)}{1+(m-1)s^*(m)} m}_{\text{group performance of } m} \tag{1.9}$$

$$E[P_{Concealment}(m)] = \sum_{m=1}^{n} \underbrace{\frac{n!}{m!(n-m)!} q^m (1-q)^{n-m}}_{\text{probability for group size } m} \times \underbrace{\frac{e^*(q,n)}{1+(m-1)s^*(q,n)} m}_{\text{group performance for m}} \tag{1.10}$$

Figure 1.5 compares the expected group performance between the disclosure policies. It shows that when the probability of being alone is high enough, expected performances are higher under concealment compared to disclosure. When the probability of being alone ($m = 1$) gets smaller (higher n and/ or higher q), the expected group performance is roughly the same across the disclosure policies. More specifically, the performance differences become less than 1, when the probability of being the only contestant is smaller than 1%. This is because being the only contestant ($m = 1$) leads to a particularly high performance under concealment compared to zero performance under disclosure. Therefore, I hypothesize:

Figure 1.5: Equilibrium expected group performance per disclosure policy



Note: The figure shows the equilibrium expected group performance (sum of individual performances) conditional on the disclosure policy for low (left panel) and high (right panel) enter probabilities. The prize is $W = 200$.

the marginal returns of the received sabotage (see Appendix A.1.6).

Hypothesis 3.1. *Concealing the number of contestants increases the expected group performance compared to disclosure when the probability of being the only contestant is not too low (at least 6%, treatments 3L, 5L, 3H).*

Hypothesis 3.2. *For a low enough probability of being the only contestant (0.4%, Treatment 5H), there is no substantial difference in the expected group performance between disclosure and concealment.*

1.3 Experimental Design

In this section, I describe the experimental design.[20] Following the model in Section 3.2, the main part of the experiment consists of a Tullock contest with exogenous enter probabilities. Subjects are part of a fixed group of potential contestants with size n and each of them becomes active with the same enter probability q. The value of the prize is worth EUR 18, so the contest is highly incentivized.

I exogenously vary the disclosure policy within subjects (full disclosure of the number of contestants m vs. full concealment), meaning that every subject makes decisions under both disclosure policies. At the same time, I vary the enter probability (low $q = 0.25$ vs. high $q = 0.75$) and number of potential contestants (small $n = 3$ vs. large $n = 5$) between subjects to study both disclosure rules under different scenarios. In this way, I vary the probability of being the only active contestant ($\mathbb{P}[m = 1] \in \{0.004, 0.06, 0.32, 0.56\}$) and also study the comparative statics of group size uncertainty. Lastly, under group size disclosure, subjects make several decisions conditional on all possible realized group sizes m, which allows me to study the comparative statics of different known group sizes m. For an overview of the experimental conditions, see Figure 1.1.

The main part of the experiment consists of 35 rounds of the contest. To ensure incentive-compatibility of each single round, I pay the average of 3 randomly determined rounds only.[21] These randomly chosen payments are displayed on the last page of the experiment only. Depending on the treatment, participants are assigned to a corresponding fixed group of 3 or 5. They stay in that group until the end of the main part and only interact with other participants of this group. Therefore, I can treat each group as a statistically independent observation. Additionally, the provided feedback of the other group members is not tied to their identities but is presented anonymously in a randomized order to reduce dynamic effects such as retaliation, reputation building, and tacit collusion across rounds.

[20] The experiment received ethical approval from the Ethics Committee of the University of Mannheim.

[21] See Azrieli et al. (2018a) for a theoretical discussion on incentive compatibility. I decided to pay the average of three rounds instead of one single round, to contribute to the maintenance of a more reliable and satisfied subject pool. Empirically, I do not observe any last-round effects.

To reduce experimenter demand and priming effects, the instructions are held on an abstract level, without using the words ‘effort’, ‘sabotage’, ‘contest’, or ‘opponents’. Instead, I call effort ‘Option A’ and sabotage ‘Option B’. Without this framing, both choices are simply tools to increase own winning probabilities with different marginal returns. Therefore, to capture the value-creating effects of effort and the value-destroying effects of sabotage, I incentivize the resulting sum of individual performances (group performance), which is positively affected by effort and negatively by sabotage. Specifically, to incorporate these value-creating and value-destroying externalities, I include donations to a charity that depend on the group performance.[22] In this way, when players exert effort, they increase their winning probabilities and the donations, and when they exert sabotage, they increase their winning probabilities but at the additional cost of decreasing the donations.[23] Note that the equilibrium predictions are not influenced by the inclusion of donations, as they do not influence the individual payoffs. Additionally, even if contestants have a preference for donations, effort and sabotage levels are only marginally different, and the comparative statics remain unchanged (see Appendix A.1.7).[24]

Figure 1.6 depicts an overview of the experimental design. The experiment starts with an extensive Tutorial and is followed by section 1. Section 1 contains the main part of the experiment, where Part A is designed to study decisions under group size disclosure and the comparative statics about the influence of a known realized group size m. Part B is designed to study decisions under group size concealment and the comparative statics of the influence of the number of potential contestants n and their enter probabilities q. Part C is identical to Part A. By comparing the choices of Part B to the choices of Part A and C, I compare the effects of the disclosure policies. Part A is repeated 15 times, followed by 15 repetitions of Part B, followed by 5 rounds of part C. The reason why I repeat another 5 rounds of group size disclosure in part C is to control for potential order effects.[25] In section 2, I elicit social value orientation (SVO), spiteful preferences, risk, loss, and ambiguity aversion, and

[22]Former experimental literature on sabotage in contests includes a principal in their experiment whose payoff is determined by the performance of the contestants (Harbring and Irlenbusch, 2011, 2008). While this procedure requires an additional participant per group, the same goal can be achieved by including donations to a charity.

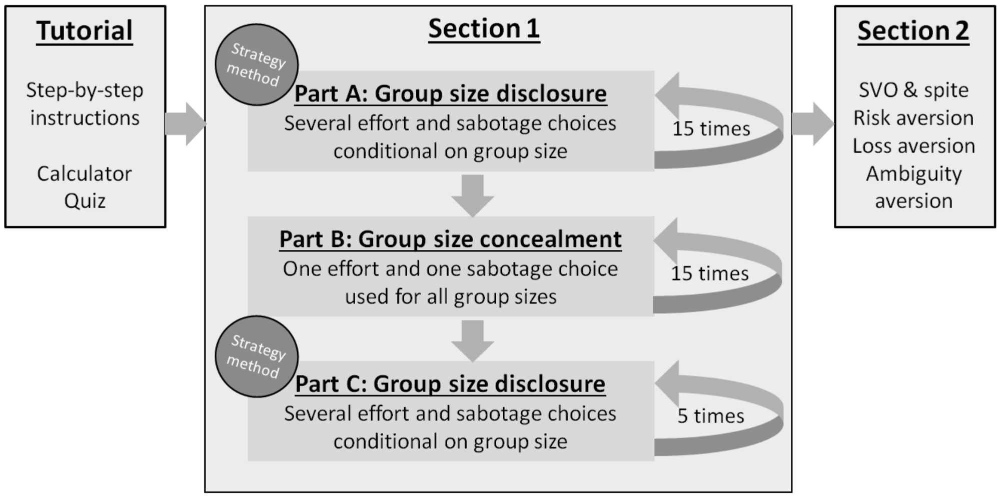
[23]The donations are calculated as follows: $donations = \sum_{i=1}^{m} y_i + 10$, where m is the number of all active players and y_i the individual performance of player $i \in M$.

[24]To eliminate heterogeneous preferences for specific charities across participants, I include five charities from various sectors (Amnesty International, Doctors Without Borders, German Red Cross, Greenpeace, and UNICEF). After all sessions were conducted, one of the charities was randomly selected for all groups. Subjects were instructed about the random selection of one charity.

[25]Appendix A.3.1 shows a small negative time trend over all rounds. The results, however, are not impacted by the time trend. Specifically, the impact of the disclosure policy is very similar between the change from disclosure to concealment in round 16 and from concealment to disclosure in round 31. Additionally, the comparative statics of disclosure and concealment are not impacted by the slight time trend (see Appendix A.3.2 and Appendix A.3.3).

standard demographics.

Figure 1.6: Experimental design



I now describe the experimental procedure in detail (see Appendix A.4 for the experimental instructions). To make sure that participants understood the experiment, they started with an extensive tutorial. In this tutorial, the rules were explained carefully and subjects could make practice choices with the computer making random choices for their opponents. The tutorial started with a simple contest scenario and successively added layers to facilitate understanding. At the end of the tutorial, participants had to answer comprehension questions to ensure understanding and could only proceed until they answered all of them correctly. During the tutorial and throughout section 1, participants had access to a probability calculator, where they could try out different effort and sabotage levels (see Figure A.8 in Appendix A.2).[26] As the contest's prize was EUR 18, participants had high incentives to work through the tutorial thoroughly and were given many tools to understand the game properly.

After the tutorial, participants started with Part A. Figure 1.7 depicts the elicitation procedure of Part A. In each round of Part A, subjects received an endowment of 200 points[27] and could use this to invest in effort ('Option A') and sabotage ('Option B').[28] They were

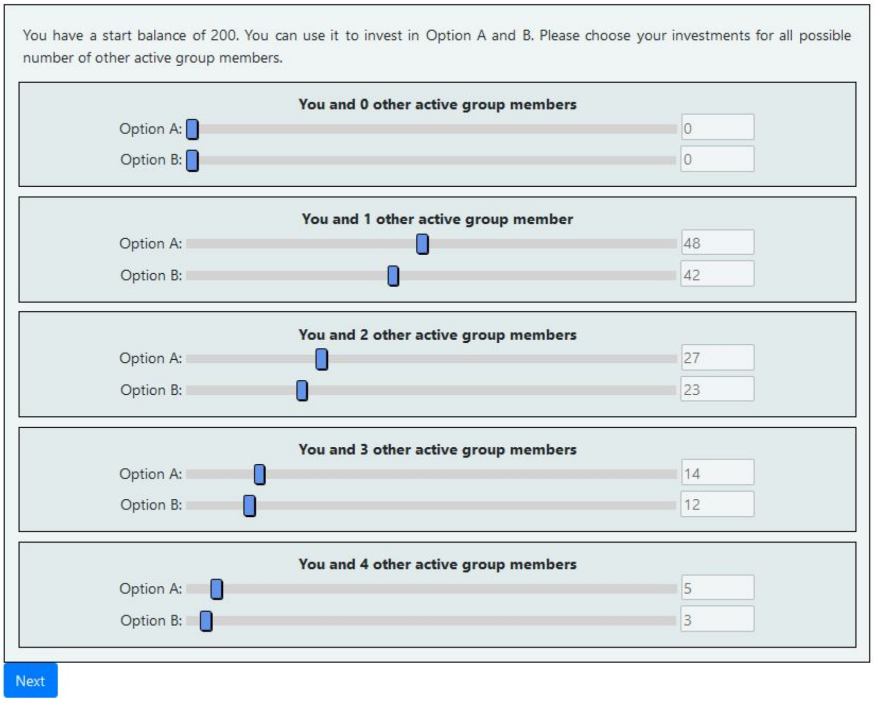
[26]Participants could enter their own levels of effort and sabotage and do the same for all other active participants. In the simplified version, the calculator assumed all others to make the same decision. Subjects could switch to the advanced version, where they could indicate different choices for every other active participant. The probability calculator then dynamically showed them their winning probabilities for all possible group size realizations with dynamic pie charts. Additionally, the donations for the specific group sizes were shown, as well as their payoffs conditional on winning or losing.

[27]I used an experimental currency called 'points' with an exchange rate of 100 points = EUR 9.

[28]Using a chosen effort and sabotage design goes in line with (e.g. Harbring and Irlenbusch, 2011, 2008) and allows me to more cleanly test the theoretical predictions. For instance, effort provision in real-effort tasks seems to be insensitive to monetary incentives (Erkal et al., 2018).

asked for their choices for all possible realized group sizes prior to their realization.

Figure 1.7: Effort and sabotage elicitation under group size disclosure (Part A & Part C)



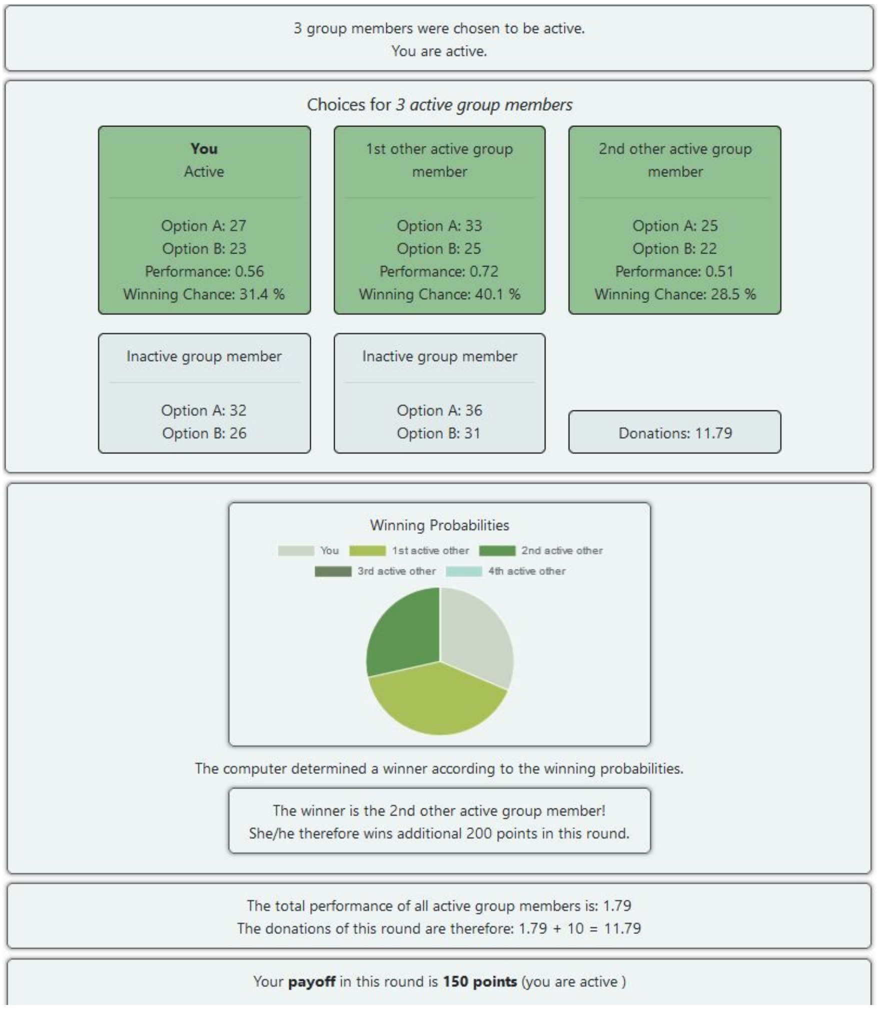
After all group members made their choices, the contest was realized as follows (see Figure 1.8): First, the computer decided who became active according to the enter probabilities.[29] After that, the computer calculated their performances and winning probabilities with the choices for the specific realized group size. It then randomly determined a winner according to the winning probabilities and calculated the donations. Then, participants received feedback about all other group members' effort and sabotage levels (including from the inactive group members) as well as the performances, winning probabilities, the winner, and the group's donations. The identities of their other group members were not disclosed in the feedback, as they were called either 'other active player' or 'other non-active player' in a randomized order.[30] Additionally, the computer calculated and showed the individual payment of the round.[31] Then the next round began.

[29]If none of the participants were chosen to become active, the computer decided for everyone anew. This procedure does not influence the relevant group size probabilities conditional on being active.

[30]Including all (active and inactive) group members' effort and sabotage levels in the feedback minimizes learning effect differences between the treatments. Otherwise, as enter probabilities are different across the treatments, there would be more feedback in the *5H*, and *3H* treatments compared to the *5L*, and *3L* treatments. Additionally, the display of the order of the group members was randomized such that it was more difficult to identify another participant's dynamic decisions.

[31]If a player was chosen to be active, all costs for sabotage and effort were deducted from the endowment. If this active player won, the prize was added to the payment. Inactive players received the endowment and costs for the stated investments were not deducted.

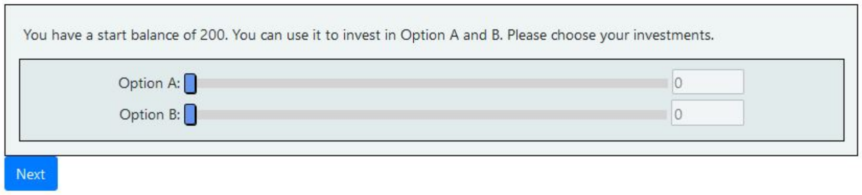
Figure 1.8: Contest realization and feedback after effort and sabotage elicitation



After finishing all 15 rounds of Part A, participants received short instructions for Part B, and went through 15 rounds of Part B. In the instructions of Part B, I communicated the group size probabilities conditional on participation instead of enter probabilities for better understanding, following Boosey et al. (2017). Participants could access these probabilities throughout the whole Part B (see Figure A.7 in Appendix A.2). In Part B, participants had to indicate one effort and one sabotage decision prior to the group size realization (see Figure 1.9). This one decision each was then taken for any group size realization. The contest realization and the feedback were the same as in Part A, with the only difference that the one effort and sabotage levels were taken for any number of active contestants. After finishing Part B, participants completed 5 additional rounds of group size disclosure in Part C.

In section 2, I used the 6-item primary scale of the SVO Slider Task (Murphy and Ackermann, 2014; Murphy et al., 2011b) to elicit prosocial preferences (see Table A.13). The

Figure 1.9: Effort and sabotage elicitation under group size concealment (Part B)



choices result in a continuous measure, the SVO-angle, which ranges from $-16.26°$ to $61.39°$. It represents a participant's prosociality, where a higher angle represents a higher prosociality. I additionally included the 3 items of the spite task to elicit spiteful preferences also used by Mill and Stäbler (2023); Mill and Morgan (2022b,a), and Kirchkamp and Mill (2021b). The spite score is calculated by dividing the destroyed points relative to the maximally possible points and hence ranges between 0 and 1. One of the 9 items was randomly determined for payment. Afterwards, I elicited risk aversion, loss aversion, and ambiguity aversion using a lottery list similar to the methods used by Holt and Laury (2002) and Sutter et al. (2013), following Boosey et al. (2017) (see tables A.14, A.15, and A.16).[32] The risk and loss aversion lists were presented in a random order, ambiguity aversion was always in third place because its elicitation builds on the risk aversion list. One row of one of the lists was chosen randomly for payment. At the very end, participants answered a questionnaire to elicit standard demographics that included age, gender, highest degree, the field of study, and a self-report of how concentrated they were and how well they understood the experiment.

The experiment was conducted online using the subject pool from the Mannheim Laboratory for Experimental Economics. Sessions were organized through Zoom meetings, where the experimenter welcomed the participants and distributed individual participation links to the software. Subjects could not turn on their microphones or videos and also could not chat with each other. Additionally, the experimenter ensured anonymity by removing the subjects' names when admitting them from the waiting room.

This online setting has several advantages. First, it ensures anonymity, and thus decreases reputational concerns, which may be especially important for sabotage decisions with negative externalities on donations. Second, it excludes social ties and peer effects, as subjects do not know, who else participates in the session. Lastly, relying on the university's subject pool may increase motivation, concentration, and accuracy in the decision-making process compared to other online samples.

[32] In each list, participants chose between a gambling lottery and a certain amount of money. Risk aversion and loss aversion are constructed with the row number, where participants switched between the gamble and the certain amount. Ambiguity aversion is constructed by taking the difference from the row number where participants switched in the risk list and the ambiguity list.

1.4 Results

In this section, I present the results of the experiment. I conducted the experiment online with the subject pool of the Mannheim Laboratory for Experimental Economics (mLab). Subjects were recruited via ORSEE (Greiner, 2015), the experiment was programmed in oTree (Chen et al., 2016), and the online sessions were implemented with Heroku servers. Overall, 196 subjects participated in the experiment.[33] The average duration was about 80 minutes and the average payoff was EUR 21.50 (min = EUR 10.56, max = EUR 32.25). The average donations per group amounted to EUR 2.83. The mean age was 23.6 years and 50% of the subjects were female.

Throughout the results section, I rely on non-parametric Wilcoxon signed-rank tests for within-subjects comparisons and on non-parametric Mann–Whitney U tests for between-subjects comparisons. The unit of analysis is the fixed groups. As everyone makes their effort and sabotage decisions conditional on being active, but prior to knowing whether they become active or not, I analyze all effort and sabotage decisions of all the participants in each round, including those who were not chosen to become active in a specific round.

I start with the main results about the differences between the disclosure policies in Section 1.4.1, and subsequently also show the comparative statics with respect to realized group sizes m under disclosure, and with respect to the number of potential contestants n and their enter probabilities q under group size concealment in Section 1.4.2.

1.4.1 Comparing Disclosure Policies

In this section, I compare the effects of disclosing the number of contestants compared to concealment. In Section 1.4.1, I find no differences in average effort, sabotage, and expected payoffs between the disclosure policies. Subsequently, in Section 1.4.1, I find that the sum of individual performances (group performance) is higher under concealment, provided that the probability of being alone is at least 6%. Given that the sum of individual performances reflects the amount of value that is induced by the contest, the designer prefers concealing the number of contestants in this case.

To compare the choices of the two policies, I compute the average expected values based on the elicited values. For this, I take the weighted sum of all elicited values over all possible group size realizations (and combinations of opponents) weighted by their probabilities. In this way, I use all the elicited choices of every player in each round. As in the theory part, I do this conditional on at least one player being active. The results thus show the

[33] I excluded one participant who dropped out due to internet problems, in accordance with the preregistration, which indicated the exclusion of subjects, who leave early or have continuous technical problems. Hence, I analyze the behavior of 195 subjects.

average expected effort, sabotage, received sabotage, payoffs, and the resulting expected group performance from a player's view conditional on being active. All results can be replicated by focusing on the actually implemented choices (see Appendix A.3.1).[34]

Furthermore, there are slight time trends in the expected effort and sabotage levels, as well as in the expected group performance (see Appendix A.3.1). Therefore, as robustness checks, first, I analyze only 5 rounds each around the changes of the disclosure policy (i.e., rounds 10-20 and 25-30) to focus on the induced differences.[35] Second, I run regressions that include the pre-registered controls.[36]

Effort, Sabotage, and Expected Payoff

Figure 1.10 shows the differences in the average expected effort, sabotage, and average expected payoff between the disclosure policies pooled over all treatments. As theory predicts (see Hypothesis 2.1), I do not find any significant difference in the average expected effort and sabotage levels across the two disclosure policies. Even though there is a marginally significant ($p < 0.1$) increase in effort under disclosure, this difference is not robust to either focusing on the subset of rounds around the change or the regression analysis, which among other variables, controls for the time trend (see Appendix A.3.1).[37] Moreover, I also do not find any significant difference in the expected individual payoffs, as predicted (see Hypothesis 2.2).[38] All robustness checks do not find any significant difference (see Appendix A.3.1).

The insensitivity of the exerted effort, sabotage, and the resulting expected payoffs towards the disclosure policy does not depend on the specific setting, as I do not find differences in effort, sabotage, or expected payoffs between the disclosure rules in any of the treatments individually (see Appendix A.3.1). The only exceptions are sabotage levels in Treatment *3L*, which are slightly higher under concealment($p < 0.05$ in the robustness checks, otherwise $p < 0.1$). Summarizing, I find the following:

[34] The realized values do not rely on all elicited values, but only on the random draw of active contestants in each round and their elicited values for the randomly realized group size and therefore add noise in each round.

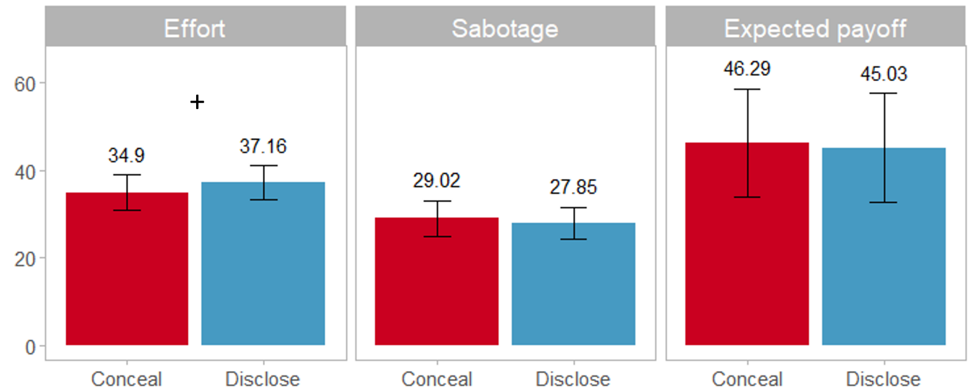
[35] I did not specify this robustness check in the pre-analysis. However, it is consistent with analyzing this subset of rounds around the policy changes and controlling for time effects.

[36] The controls are: Being active in the round before, having won in the round before, average sabotage and effort levels of other participants in the rounds before, round, the treatments, realized group size in the round before, how often won in the rounds before, SVO, spite, risk, loss and ambiguity aversion, age, gender, highest degree, the field of study, the degree of concentration and understanding.

[37] Instead, the regression analysis shows a significant ($p < 0.05$) positive increase in sabotage under concealment. However, a Cohen's D of -0.09 for sabotage shows that even if there are significant differences between the disclosure policies, this difference can not be considered to be very substantive. Additionally, in no other robustness check do I find this significant increase.

[38] The expected payoffs exclude the 200-point endowment in each row and thus represent the expected payoff from the contest.

Figure 1.10: Results average expected effort, sabotage, and payoffs per disclosure policy



Note: The figure shows average expected effort, sabotage, and payoffs conditional on the disclosure policy, pooled over all treatments. Error bars show 95% confidence intervals. Significance levels: + $p < 0.10$

Result 1.1. *Concealing the number of contestants does not significantly change average expected effort and sabotage levels, except when the probability of being alone is high (52%, Treatment 3L), concealment leads to higher sabotage levels.*

Result 1.2. *The average expected payoff does not significantly differ between concealing and disclosing the number of contestants.*

Group Performance

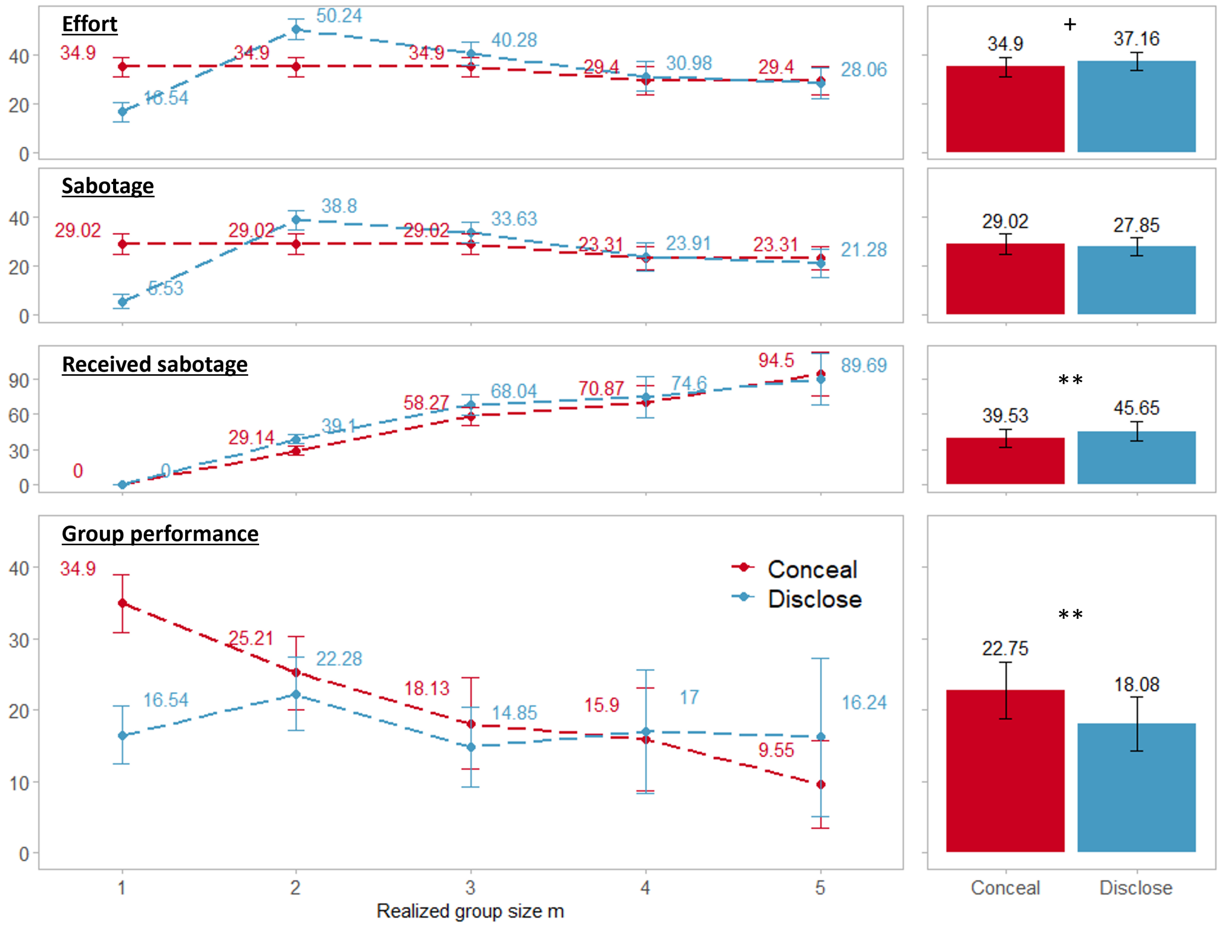
So far, I showed that subjects do not significantly change their average effort and sabotage levels between the disclosure policies and, as a consequence, their expected payoffs do not differ. Hence, they are ex-ante indifferent towards the chosen disclosure policy. From a welfare point of view, theory predicts that concealment leads to a higher sum of individual performances and hence to more created value. In this section, I study whether the experiment shows that group performances are indeed higher under concealment.

Figure 1.11 overall illustrates how the group performance is shaped, pooled over all treatments.[39] The group performance is defined as: $\sum_{i\in M} y_i = \sum_{i\in M} \frac{e_i}{1+\sum_{j\in M_i} s_j}$, and rows 1 and 2 show the average effort and sabotage levels (e_i and s_i), row 3 the average received sabotage ($\sum_{j\in M_i} s_j$), and row 4 the average group performance. The panels on the LHS show these values depending on the realized group size, whereas the panels on the RHS show the weighted averages.

The bar chart on the RHS of row 4 shows that concealing the number of contestants significantly increases group performance ($p < 0.01$). This result can be replicated in both

[39] The data for the realized group size of $m = 4$ and $m = 5$ come from treatments *5L* and *5H* only. Comparative statics look relatively similar across treatments.

Figure 1.11: Results group performance



Note: The right panels show differences between the two disclosure policies in the expected effort and sabotage levels, the received sabotage, and the resulting group performance. The left panels show them conditional on the realized group size. Error bars show 95% confidence intervals. Significance levels: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

robustness checks ($p < 0.05$ and $p < 0.1$) (see Appendix A.3.1). The panel on the LHS of row 4 depicts those differences depending on the realized group size m. It shows that the group performance difference between the disclosure policies is primarily driven by the case when contestants do not face any competitors ($m = 1$). This goes in line with theory because, under concealment, subjects have to choose one effort level without knowing the group size and hence exert a large amount of effort even if they end up being the only contestant and win the contest with certainty. If they know that they are the only contestant, they exert much less effort.[40] The important factor that induces performance differences is the combination of the exerted effort and the received sabotage depending on the realized group size. Specifically, when a contestant does not face any competitor, she is not subjected to any sabotage,

[40]They still provide non-zero effort because effort creates value and increases the donations in the experiment.

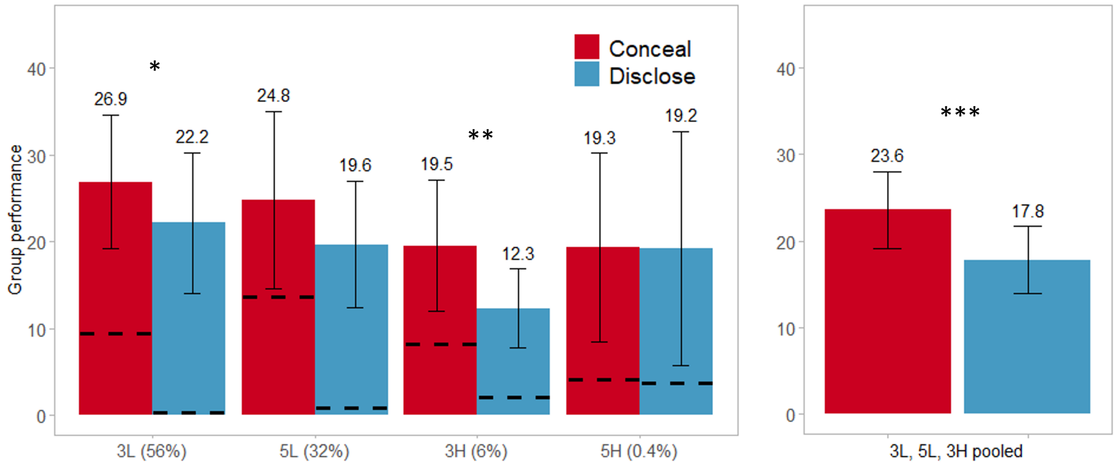
simply because there are no others who sabotage her. Because of this, the substantive effort difference between the disclosure rules when alone ($m = 1$), translates directly into a large performance difference. For all other realized group sizes, contestants do receive sabotage and hence, even if there are effort differences, the resulting group performances are not significantly different because of the sabotage that they receive from each other. Specifically, for realized group sizes of $m = 2$ and $m = 3$, subjects exert significantly ($p < 0.01$) higher effort under disclosure (row 1), yet, also receive significantly ($p < 0.01$) higher levels of sabotage (row 3), leading to not significantly different group performances. For realized group sizes of $m = 4$ and $m = 5$, there are no significant differences in the exerted effort, received sabotage, and resulting group performances between the disclosure policies.[41]

Even though I find a significant difference between the disclosure policies, the difference is not as pronounced as predicted. There are two reasons for this. First, subjects on average provide a substantial amount of effort, even when they know that they are the only contestant. This is because effort is constructive and increases the donations. Second, for all other realized group sizes ($m > 2$), group performances are overall much higher than predicted under both policies (see Figure 1.4), reducing the effect of the difference for a realized group size of one. This is because of substantial heterogeneity in the exerted effort and sabotage between group members. The group member, who exerts the most effort, on average receives the least sabotage and thus maintains a higher performance (see Appendix A.3.1). Despite these heterogeneities, I still find the predicted increase in group performance under concealment.

Next, I test my theoretical prediction (see Hypothesis 3.1 and Hypothesis 3.2) that concealment leads to higher group performances only when the probability of being the only contestant is roughly larger than 1% (56% in Treatment *3L*, 32% in Treatment *5L*, and 6% in Treatment *3H* opposed to 0.4% in Treatment *5H*). Figure 1.12 depicts the average group performance per treatment conditional on the disclosure policy in comparison to the Nash equilibrium predictions (dashed black lines). It shows a significant increase in the group performance under concealment for *3L* ($p < 0.05$) and *3H* ($p < 0.01$) and a non-significant increase for *5L*. For *5H*, there is no significant difference between the disclosure policies, as predicted. Moreover, because theory predicts an increase for treatments *3L*, *5L*, and *3H*, I pool them and find a significant increase in group performances under concealment ($p < 0.001$). This difference is quite substantial with an increase under concealment compared to disclosure of around 30%. The robustness checks (see Appendix A.3.1) confirm the significant differences in all cases but for Treatment *3H*, where I do not find any significant differences, yet the increase is qualitatively replicated.

[41] Only the group performance difference for a realized group size of 5 is marginally significant ($p = 0.09$).

Figure 1.12: Results group performance per treatment



Note: The bar charts show the average group performance conditional on the disclosure policy and on treatments. Percentages in parentheses show the probability of being the only contestant in each of the treatments. Black dashed lines show the Nash equilibrium predictions. The error bars show 95% confidence intervals. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Result 2.1. *When the probability of being alone in the contest is at least 6% (3H, 3L, 5L), concealment leads to higher group performance.*

Result 2.2. *When the probability of being alone in the contest is 0.4% (5H), I do not find any differences between the disclosure policies.*

1.4.2 Comparative Statics under Disclosure and Concealment

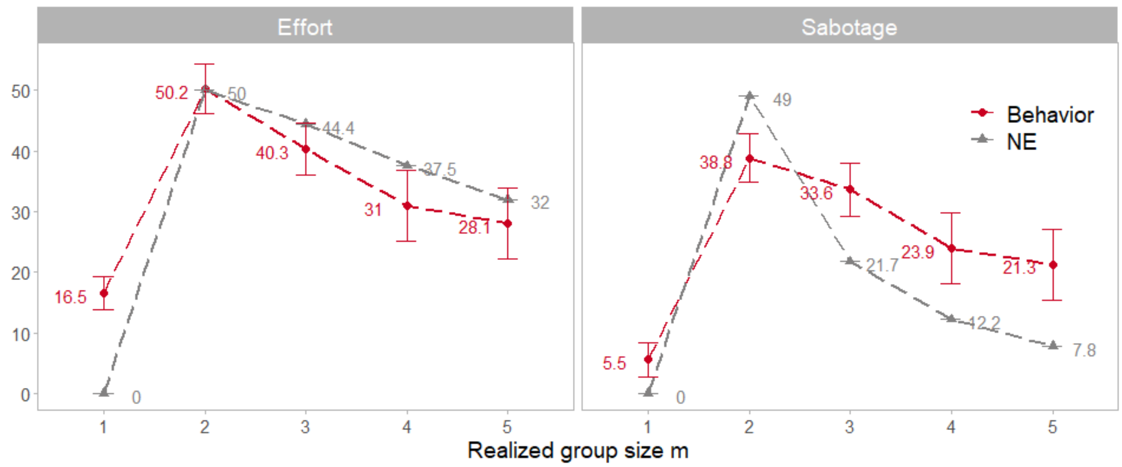
In this section, I study how different known group sizes m influence effort and sabotage levels in Section 1.4.2 and how the number of potential contestants n and their enter probabilities influence effort and sabotage levels under group size uncertainty in Section 1.4.2.

Known Group Sizes (Group Size Disclosure)

Figure 1.13 depicts mean effort and sabotage levels under group size disclosure conditional on the realized group size compared to the Nash equilibrium predictions. Averages are computed over all rounds (of Part A and Part C) and pooled over all treatments. The figure suggests that effort levels follow very closely the equilibrium predictions. Specifically, I find a significant decrease in effort from a realized group size of 2 to 5 ($p < 0.001$). This decrease in effort is in line with the experimental contest literature without sabotage (Anderson and Stafford, 2003; Sheremeta, 2011; Morgan et al., 2012; Aycinena and Rentschler, 2019).

Furthermore, I find that sabotage also decreases significantly from $m = 2$ to $m = 5$ ($p < 0.001$), as predicted by theory (see also (Konrad, 2000)). A regression analysis reveals

Figure 1.13: Results average effort and sabotage levels per realized group size



Note: The figure illustrates effort and sabotage levels under group size disclosure as a function of the realized group size. Grey lines show the equilibrium predictions. Red lines depict the averages of the elicited behavior of the experiment pooled over all treatments. The error bars show 95% confidence intervals.

a significant ($p < 0.001$) negative effect of a realized group size on effort and sabotage for group sizes $m > 1$ (see Appendix A.3.2).

The decrease in sabotage is slightly less steep than predicted, leading to over-sabotage for larger group sizes. In particular, sabotage levels are significantly ($p < 0.001$) below the Nash equilibrium for a group size of 2, and significantly ($p < 0.001$) above for the group sizes of 3, 4, and 5. In the experimental contest literature, it is common that subjects over-invest in effort compared to the Nash equilibrium (Sheremeta, 2018; Dechenaux et al., 2015; Sheremeta, 2013).[42] As the total amount of exerted sabotage is added up over the number of active contestants, this over-sabotage is particularly harmful, as it leads to more destroyed value for larger realized group sizes.

All results remain robust when analyzing different pre-registered sets of subrounds (see Appendix A.3.2) and when running a regression analysis (see Appendix A.3.2).[43] To summarize, I find the following:

Result 3.1. *An increase in the group size (for $m > 1$) decreases effort and sabotage levels.*

Result 3.2. *There are above-equilibrium sabotage levels for realized group sizes larger than 2.*

I propose two concepts that can explain the increases in over-sabotage in the group size. First, a modified version of joy of winning – constant winning aspiration – postulates

[42]There is no overbidding in effort and no joint overbidding when aggregating both effort and sabotage levels. The sum of effort and sabotage levels is not significantly higher than the sum of the Nash equilibrium predictions: $m = 3$: $p = 0.078$, $m = 4 : p = 0.430$, $m = 5 : p = 0.114$.

[43]Additionally, the comparative statics remain stable over time (see Appendix A.3.2).

that joy of winning increases linearly with group size Boosey et al. (2017). Hence, subjects experience greater joy, when they win against more competitors, which makes them over-invest for larger group sizes. This cannot explain, however, why there exists over-sabotage but not an overexertion of effort.

Second, if subjects have spiteful preferences (e.g. Levine, 1998; Morgan et al., 2003a), they receive additional utility for harming others. As sabotage's harm increases in the number of competitors, spite's utility gains also increase in the number of competitors and thus can explain the above-equilibrium sabotage levels for larger group sizes. This explanation is supported by the significant positive correlation between spiteful preferences and sabotage (see Table A.9 in Appendix A.3.2).

Group Size Uncertainty

I now turn to the case of group size uncertainty. Figure 1.14 depicts average effort and sabotage levels conditional on the enter probability (high vs. low) and the number of potential contestants. For high entering probabilities ($q = 0.75$), I find a significant ($p < 0.01$) decrease in effort and sabotage levels when the number of potential contestants increases from 3 to 5, as predicted. This part of the comparative statics is therefore supported by the evidence. In the case of low enter probabilities ($q = 0.25$), however, I do not find the predicted increase in effort and sabotage levels. Instead, I find a slight (non-significant) decrease in effort and sabotage. This deviation from theory does not come from distortions related to the uncertainty of the group size, as subjects take the weighted average of their known group size choices (see Section 1.4.1). This result complements the results of Boosey et al. (2017) who find a significant increase in effort levels for an increasing group size for low entering probabilities but no significant increase for high entering probabilities.

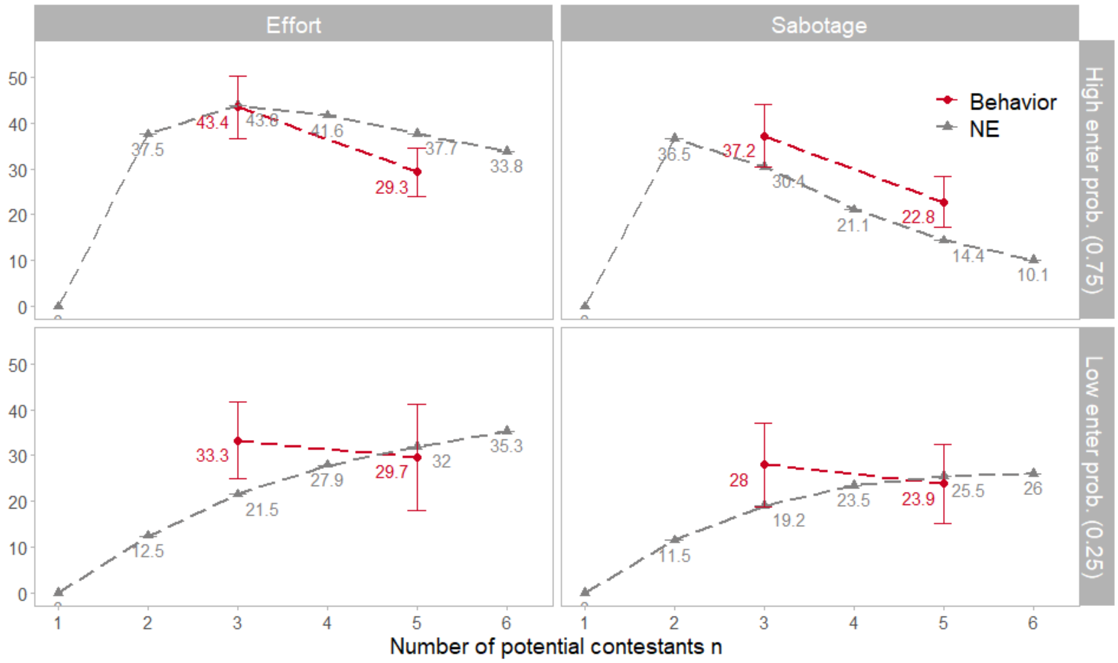
The results remain robust to analyzing only the pre-registered subrounds and running a regression analysis with the pre-registered controls (see Appendix A.3.3 and Appendix A.3.3).[44] Summarizing, I find:

Result 4.1. *For high enter probabilities ($q = 0.75$), an increase in the group size (from $n = 3$ to $n = 5$) decreases effort and sabotage levels.*

Result 4.2. *For low enter probabilities ($q = 0.25$), an increase in the group size (from $n = 3$ to $n = 5$) does not significantly change effort and sabotage levels.*

[44]Specifically, I find a significant decrease (at least $p < 0.05$) in effort and sabotage levels from $3H$ to $5H$ for the subset of rounds 1-7 and rounds 8-15 and in the regression analysis. For the single round 1, I only qualitatively replicate the decrease ($p = 0.1994$). I do not find any significant differences between the treatments 3L and 5L in any of the robustness checks. Moreover, I find a slight decrease in effort and sabotage over time, however, the comparative statics remain stable (see Appendix A.3.3).

Figure 1.14: Results effort and sabotage under group size uncertainty



Note: The figure illustrates effort (left panels) and sabotage (right panels) decisions under group size uncertainty (Part B) for high (upper panels) and low (lower panels) enter probabilities. The x-axes show the number of potential contestants n. Grey lines depict the equilibrium predictions and red lines the averages of the elicited choices. Error bars show 95% confidence intervals.

1.5 Discussion and Conclusion

In this paper, I provide a theoretical and experimental analysis of how disclosing the number of contestants affects effort and sabotage levels compared to concealing this information. Since contests are often used to increase the productivity of workers or companies, I compare the resulting differences in the created value. To do so, I compare the resulting sum of individual performances (group performance) because it incorporates the productive effects of effort on their own performance and the destructive effects of sabotage on others' performances.

I model sabotage in a Tullock contest with exogenous enter probabilities, where the designer commits to either always conceal or disclose the number of contestants. According to the theoretical analysis, this decision should not affect average effort, sabotage, or expected payoffs. This is because when agents do not know the number of competitors, their equilibrium levels are the weighted sum of their choices for those specific group sizes. The choice of the disclosure policy does, however, induce differences in the resulting group performance. This is because performances depend on the combination of effort and sabotage, and have decreasing marginal returns in the received sabotage. When agents do not know the realized group size, they provide one effort and one sabotage level for all realized group sizes. In contrast, when they know the realized group size, they can adjust their effort and sabotage levels to the number of contestants. As a consequence, the distribution of effort and sabotage differs between the disclosure policies depending on the realized group size and hence leads to performance differences. An essential feature of group size uncertainty is the probability of being the only contestant. When contestants know that they are alone, they do not provide any effort because they know that they will win with certainty, leading to a performance of zero. In contrast, when they do not have this information, they provide a substantial amount of effort, even when they are the only contestant. Additionally, they do not receive any sabotage in this case and hence the resulting performance is particularly pronounced. For all other realized group sizes, performance differences are relatively small between the policies, leading to a higher performance under concealment.

This result demonstrates that it is important to consider sabotage when comparing group size disclosure policies, as omitting the possibility of sabotage can lead to wrong conclusions. Indeed, in a standard contest with symmetric agents and linear costs but without sabotage, there are no differences between the disclosure policy (Lim and Matros, 2009). By incorporating the welfare effects of sabotage, I show that the choice of the disclosure policy matters. This adds to other, but more specific theoretical settings, where differences between the two policies arise.[45]

I run an experiment to test my theoretical predictions. In line with theory, the first key result is that a designer can increase the sum of individual performances, and thus the amount of created value, by concealing the number of contestants. However, this only works if the probability of being the only contestant is not too low. This is because, as predicted, the difference in group performance is driven by substantially higher performances under concealment when contestants do not face any competitors. For all other realized group sizes, I find no significant difference in the resulting group performance. Therefore, I do not find any difference in group performances, when the probability of being alone is negligible (0.4%) but otherwise (at least 6%), I find that concealment leads to higher group performances. With this result, I provide experimental evidence that from a welfare perspective the possibility of sabotage leads to differences between the disclosure policies. Consequently, the experiment can be seen as a successful robustness check to the theory by allowing for heterogeneities in prize valuations, moral costs, degrees of sophistication, risk aversion, probability distortion, and other heterogeneities among contestants. The experiment thus extends the

[45]These settings include different prize valuations together with different enter probabilities or endogenous entry (Fu et al., 2016; Chen et al., 2023), convex or concave cost structures (Jiao et al., 2022; Chen et al., 2017), a strictly convex or concave characteristic function of the Tullock contest (Feng and Lu, 2016; Fu et al., 2016), or the existence of bid caps (Wang and Liu, 2023; Chen et al., 2020a).

generalizability of the theoretical finding.

With this first key finding, I provide evidence that sabotage matters when considering the welfare effects of a group-size disclosure policy. Contrary to the experimental findings of Boosey et al. (2020) and Jiao et al. (2022), who find that concealment leads to lower effort provisions, I find that when subjects can sabotage each other, concealment leads to an increase in performance. I also add to the sabotage literature, by providing a way how to mitigate the welfare-destroying effects of sabotage. This is different from the approach of the sabotage literature that mostly discusses ways of how to decrease sabotage altogether (see Chowdhury and Gürtler (2015)).[46]

The second key result is that contestants are ex-ante indifferent between the two disclosure policies.[47] This is because I do not find any difference in average effort, sabotage, or expected payoffs between the disclosure policies, as predicted. The only exception is when the probability of being the only contestant is high (56%), where I find that concealment leads to a slight increase in sabotage.

Not finding any significant difference in effort and sabotage levels between the disclosure policies, goes in line with the experimental literature, which also does not find differences in the average effort levels in contests in most settings (Jiao et al., 2022; Boosey et al., 2020; Aycinena and Rentschler, 2019). However, when the outside option is high and enter choice is endogenous (Boosey et al., 2020), or when the cost structure is concave (Jiao et al., 2022), disclosing the number of contestants can induce higher efforts. I provide the special case of a high probability of being the only contestant (56%), where disclosure leads to lower sabotage, but not to a difference in effort. In all other cases, I find that under concealment, choices are the weighted sum of subjects' choices under disclosure. Consequently, subjects in my experiments do not seem to exhibit probability distortions. This is different from the experiment of Boosey et al. (2017), where the authors explain their observed effort levels under group size uncertainty with probability distortions.

The practical implication for a designer is that she can induce higher performances by concealing the number of contestants without curbing their productive efforts or expected payoffs. This is a notable result because the created value can be increased without requiring contestants to exert additional effort. Rather, it enhances the productivity of the exerted effort by mitigating the destructive effects of potential sabotage. Furthermore, con-

[46]Possible such ways include reducing the prize spread (Harbring and Irlenbusch, 2011, 2005; Del Corral et al., 2010; Vandegrift and Yavas, 2010; Lazear, 1989), increasing the number of contestants (Konrad, 2000), increasing the costs for sabotage (Balafoutas et al., 2012), and other information disclosure policies such as concealing intermediate rank information (Charness et al., 2014; Gürtler et al., 2013; Gürtler and Münster, 2010), or revealing the identity of the saboteur (Harbring et al., 2007).

[47]If they do not have preferences over the produced value. Otherwise, they would prefer concealment when the probability of being the only contestant is not too low.

cealing the number of contestants is an easy-to-implement tool because not disclosing the number of contestants simply requires deliberately omitting information about the group size. Whether concealment should be implemented, however, depends on the specific setting. This is because concealment is effective only when the probability of a player being the only contestant is not too low (larger than 6%). At the same time, if this probability is too high (56%), concealment can lead to higher sabotage levels. A designer should therefore carefully counterbalance the effects of a specific setting.

As additional results, I find evidence for the comparative statics of known group sizes under group size disclosure. When contestants know the number of competitors, a higher group size decreases effort and sabotage levels. This provides evidence for the theoretical sabotage results of Konrad (2000), which has been pointed out to lack empirical evidence (Piest and Schreck, 2021; Chowdhury and Gürtler, 2015). I also observe significant above-equilibrium sabotage levels for realized group sizes of 3, 4, and 5. This behavior can be explained by a modified version of joy of winning (constant winning aspirations), where the experienced joy increases in the number of outperformed competitors (Boosey et al., 2017), or by spiteful preferences (Morgan et al., 2003a), where agents receive utility by harming others, and hence this utility increases in the number of harmed competitors. Empirically, I find an overall positive correlation of spite with sabotage, which suggests that the observed over-sabotage is, at least, partially driven by spiteful preferences. The importance of spiteful preferences adds to other literature which shows that spiteful preferences matter in competitive settings (Mill and Stäbler, 2023; Mill and Morgan, 2022a; Mill, 2017a). Observing this significant overbidding in sabotage for larger group sizes goes in contrast to Boosey et al. (2017), who do not find any overbidding in effort levels in a contest with group size uncertainty, but in line with the experimental literature on contests with known group sizes, which consistently finds overbidding in effort (Sheremeta, 2018; Dechenaux et al., 2015; Sheremeta, 2013). Yet, I also do not find any overbidding in effort, nor in the joint effort and sabotage levels.

From a welfare perspective, observing higher-than-equilibrium sabotage and at the same time, not higher effort levels is bad news, especially for larger groups. The individually exerted over-sabotage leads to a drastic increase in the received sabotage when the number of sabotage-exerting contestants increases. Hence, more value is destroyed and individual performances diminished. This illustrates that contrary to theory, sabotage is not necessarily a 'small number phenomena' (Konrad, 2000), but sabotage is especially harmful when the group sizes become larger. As a consequence, increasing the number of contestants to decrease sabotage does not seem to be an apt tool. Instead, if a designer can set and reveal the number of competitors, she should rather determine a smaller number of competitors, as less value is destroyed.

Another additional result is that when contestants do not know the realized number of contestants, I find that an increase in the number of potential contestants decreases effort and sabotage levels when enter probabilities are high ($q = 0.75$), as predicted. When enter probabilities are low ($q = 0.25$), however, I do not find evidence for the theoretical decrease in effort and sabotage. These results complement Boosey et al. (2017) who do not find a significant difference in effort levels when enter probabilities are low, but a significant increase in effort when enter probabilities are high.

The study comes with certain limitations. For instance, I abstract from any spillovers from the exerted sabotage on a baseline productivity of all potential contestants, including non-active players. If the sabotage activities also harm all sabotaged players' baseline productivities, concealment increases this harm, as subjects exert sabotage even when they are alone in the contest. How much harm is done, and which of the counteracting forces between the disclosure policies prevails, depends on the specific parametrization of the baseline productivity and the effectiveness of sabotage on the baseline productivity. For a small enough baseline productivity or small enough effectiveness of sabotage, the results of this paper still hold. Furthermore, I assume exogenous enter probabilities, whereas many times entry into a contest may be an endogenous choice. In this paper, exogenous entry probabilities can be thought of as fixing enter beliefs and significantly reducing complexity for subjects. Endogenizing enter probabilities provides an interesting avenue for future research. This is because the possibility of exerting sabotage and getting sabotaged may attract in particular spiteful and tough players, which may potentially lead to additional differences between the disclosure policies and to even more over-sabotage for larger group sizes.

Future work should study group sizes larger than five to explore whether the behavioral pattern of over-sabotage further increases. Additionally, it would be interesting to expand the disclosure policy to not only disclosing the number of contestants but also to revealing contestants' identities. If contestants know the identities of their competitors, their sabotage activities can be better targeted. In this way, their sabotage becomes more effective, making it more beneficial to engage in such destructive behavior, leading to more sabotage overall under disclosure compared to concealment. This would further increase the benefits of not disclosing the number of competitors and underlines the welfare-enhancing effects of concealing the number of competitors. Finally, as sabotage can destroy value, future work that assesses the welfare consequences of any kind of policies should account for the possibility of such destructive behaviors. Otherwise, the welfare assessment might lead to wrong conclusions.

Acknowledgments for Chapter 1

I thank Henrik Orzen, Wladislaw Mill, Aleksandra Khokhlova, Mark Spills, Anna Abate Bessomo, Thomas Eisfeld, Theodore Turocy, Bernhard Ganglmair, and Subashish Chowdhury for their helpful comments. I appreciate comments from participants from the ZEW/ University of Mannheim Experimental Seminar, the CDSE Seminar, the GESS research day, the Max Planck Summer School on the Political Economy of Conflict and Redistribution 2021, the 16th RGS Doctoral Conference in Economics, the MaCCI IO Day 2023, the Enter Jamboree 2023, the 9th Annual 'Contests: Theory and Evidence' Conference 2023, the 2023 ESA World Meeting, the VfS Annual Conference 2023, the ESA's job-market candidates seminar series 2023, as well as from the Microeconomics Seminar at the University Carlos III of Madrid 2023, the Experimental Seminar at the University of Innsbruck 2023, and from the ECONtribute and C-SEB Design & Behavior Seminar at the University of Cologne.

Appendix A

Appendix to Chapter 1

A.1 Theory Appendix

A.1.1 First and Second Order Conditions for Maximization Problem

In order to derive the equilibrium effort and sabotage levels, I maximize the individual payoff function with respect to effort and with respect to sabotage for contestant i without loss of generality. Equation A.1 is the first order condition with respect to effort and Equation A.2 the first order conditions with respect to sabotage.

$$\frac{\partial \pi_i}{\partial e_i} = \frac{(\frac{1}{1+\sum_{j\in M_i} s_j})\sum_{l=1}^{m}(\frac{e_l}{1+\sum_{k\in M_l} s_k}) - (\frac{e_i}{1+\sum_{j\in M_i} s_j})(\frac{1}{1+\sum_{j\in M_i} s_j})}{(\sum_{l=1}^{m}\frac{e_l}{1+\sum_{k\in M_l} s_k})^2} W - C'(e_i) = 0 \tag{A.1}$$

Next, I assume without loss of generality that player i be the m-th player.

$$\frac{\partial \pi_i}{\partial s_i} = \frac{(\frac{e_i}{(1+\sum_{j\in M_i} s_j)})(\frac{e_1}{(1+\sum_{k\in M_1} s_k)^2} + ... + \frac{e_{m-1}}{(1+\sum_{k\in M_{m-1}} s_k)^2})}{(\sum_{l=1}^{m}\frac{e_l}{1+\sum_{k\in M_l} s_k})^2} W - C'(s_i) = 0 \tag{A.2}$$

The following two second order conditions hold true $\forall e_i > 0, \forall s_i, W \geq 0, \forall i \in M$, where M is the set of all active players. This indicates that the solutions to the first order conditions are maxima:

$$\frac{\partial^2 \pi_i}{\partial e_i^2} = -2\frac{(\frac{1}{(1+\sum_{j\in M_i} s_j)^2})\sum_{j\in M_i}(\frac{e_j}{1+\sum_{k\in M_j} s_k})}{(\sum_{l=1}^{m}\frac{e_l}{1+\sum_{k\in M_l} s_k})^3}W - C''(e_i) < 0 \tag{A.3}$$

$$\begin{aligned}\frac{\partial^2 \pi_i}{\partial s_i^2} = &-2\frac{(\frac{e_i}{(1+\sum_{j\in M_i} s_j)})(\frac{e_1}{(1+\sum_{k\in M_1} s_k)^3} + ... + \frac{e_{m-1}}{(1+\sum_{k\in M_{m-1}} s_k)^3})}{(\sum_{l=1}^{m}\frac{e_l}{1+\sum_{k\in M_l} s_k})^2}W \\ &+2\frac{(\frac{e_i}{(1+\sum_{j\in M_i} s_j)})(\frac{e_1}{(1+\sum_{k\in M_1} s_k)^2} + ... + \frac{e_{m-1}}{(1+\sum_{k\in M_{m-1}} s_k)^2})^2}{(\sum_{l=1}^{m}\frac{e_l}{1+\sum_{k\in M_l} s_k})^3}W - C''(e_i) < 0\end{aligned} \tag{A.4}$$

A.1.2 Proof Proposition 1

Proof. In a symmetric equilibrium, homogeneous contestants choose the same strategies. Hence, the chosen individual effort and sabotage levels are the same for everyone: $e_i = e_{-i} = e$ and $s_i = s_{-i} = s$. As there are m active contestants, everyone receives the sabotage of $m-1$ other contestants. Therefore, the received sabotage is $(m-1)s$. Further, as I assume $C(e_i) = e_i$ and $C(s_i) = s_i$, $C'(e) = C'(s) = 1$. The first order condition with respect to effort (A.1) becomes:

$$\begin{aligned}&\frac{(\frac{1}{1+(m-1)s})(m\frac{e}{1+(m-1)s}) - (\frac{e}{1+(m-1)s})(\frac{1}{1+(m-1)s})}{m^2(\frac{e}{1+(m-1)s})^2}W = 1 \\ \iff &\frac{(\frac{1}{1+(m-1)s})(m-1)(\frac{e}{1+(m-1)s})}{m^2(\frac{e}{1+(m-1)s})^2}W = 1 \\ \iff &\frac{(m-1)}{m^2}\frac{W}{e} = 1\end{aligned}$$

$$\iff e^* = \frac{(m-1)}{m^2}W \tag{A.5}$$

Likewise, the first order condition with respect to sabotage (A.2) becomes:

$$\frac{(\frac{e}{1+(m-1)s})(m-1)(\frac{e}{(1+(m-1)s)^2})}{m^2(\frac{e}{1+(m-1)s})^2}W = 1$$
$$\iff \frac{(m-1)}{m^2}\frac{W}{(1+(m-1)s)} = 1$$
$$\iff s^* = \frac{1}{m^2}W - \frac{1}{(m-1)} \tag{A.6}$$

□

A.1.3 Proof Proposition 2

Proof. Under group size concealment, the expected profit function is as follows:

$$\mathbb{E}[\pi_i] = \sum_{M_i \in \mathcal{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|} p_i(y_i, y_{-i}, M_i)W - C_i(e_i) - C_i(s_i)$$

First, I take the first order condition of the expected profit function with respect to e_i:

$$\frac{\partial \mathbb{E}[\pi_i]}{\partial e_i} = \sum_{M_i \in \mathbb{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|} \frac{\frac{1}{1+\sum_{j \in M_i} s_j} \sum_{j \in M_i} \frac{e_j}{1+\sum_{k \in M_j} s_k}}{(\frac{e_i}{1+\sum_{j \in M_i} s_j} + \sum_{j \in M_i} \frac{e_j}{1+\sum_{k \in M_j} s_k})^2} W - C_i'(e_i) = 0$$

Next, I employ symmetry and by assumption $C'(e) = 1$. As every contestant is the same, the sum over all possible combinations of other active competitors relaxes to the binomial distribution. It describes the probabilities for each realized number of other contestants $(m-1)$ out of $n-1$ potential contestants. Hence, the equation becomes:

$$\sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-1-(m-1))!} q^{(m-1)}(1-q)^{n-1-(m-1)} \frac{\frac{(m-1)e}{(1+(m-1)s)^2}}{(m-1+1)^2 \frac{e^2}{(1+(m-1)s)^2}} W = 1$$
$$\iff \sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-1-(m-1))!} q^{(m-1)}(1-q)^{n-m} \frac{m-1}{m^2}\frac{W}{e} = 1$$
$$\iff e^* = \sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-m)!} q^{(m-1)}(1-q)^{n-m} \frac{m-1}{m^2} W \tag{A.7}$$

Subsequently, I take the first order condition with respect to s_i:

$$\frac{\partial \mathbb{E}[\pi_i]}{\partial s_i} = \sum_{M_i \in \mathbb{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|} \frac{\frac{e_i}{1+\sum_{j\in M_i} s_j} \sum_{j\in M_i} \frac{e_j}{(1+\sum_{k\in M_j} s_k)^2}}{(\frac{e_i}{1+\sum_{j\in M_i} s_j} + \sum_{j\in M_i} \frac{e_j}{1+\sum_{k\in M_j} s_k})^2} W - C_i'(s_i) = 0$$

After employing symmetry, $C'(s) = 1$, and the binomial coefficient, the equation becomes the following:

$$\sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-1-(m-1))!} q^{(m-1)}(1-q)^{n-1-(m-1)} \frac{\frac{(m-1)e^2}{(1+(m-1)s)^2}}{(m-1+1)^2 \frac{e^2}{(1+(m-1)s)^2}} W = 1$$

$$\iff \sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-m)!} q^{(m-1)}(1-q)^{n-m} \frac{m-1}{m^2} \frac{1}{1+(m-1)s} W = 1 \qquad \text{(A.8)}$$

Which does not yield a closed form solution and hence I solve it numerically. Further, the second order conditions follow immediately from equation (A.3) and (A.4) and hold such that the FOC describe the maxima. □

A.1.4 Expected Effort, Sabotage, and Payoffs

In this section, I compare the expected effort and sabotage levels between disclosure and concealment. When the group size realization is zero, i.e., when there are no contestants at all, there is no effort under both disclosure policies. Therefore, it suffices to show that the implemented effort is the same conditional on at least one player participating. Hence, conditional on at least one player participating, the expected efforts are as follows:

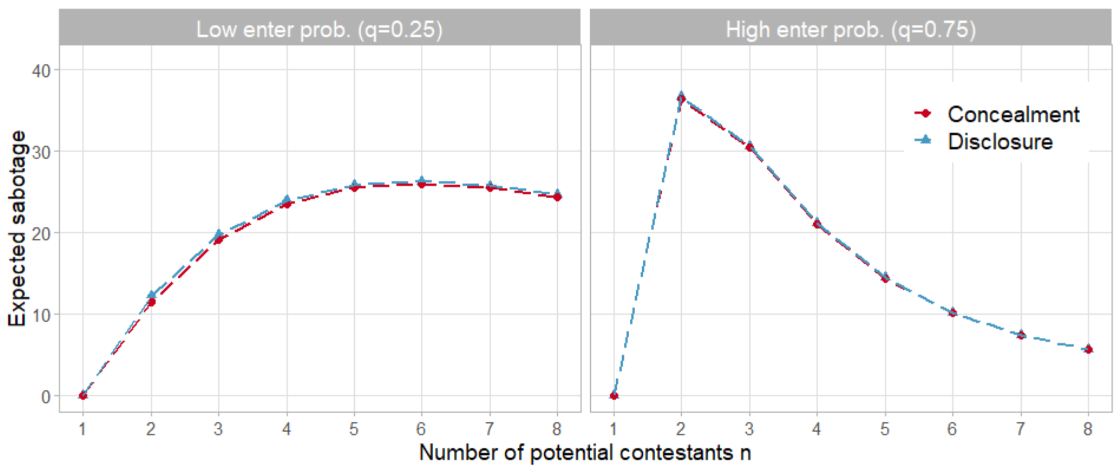
$$\textit{Concealment}: \mathbb{E}[e] = \sum_{(m-1)=0}^{n-1} \underbrace{\frac{(n-1)!}{(m-1)!(n-m)!} q^{m-1}(1-q)^{n-m}}_{\text{Probability of group size } m} \underbrace{e^*}_{\text{Effort concealment}} = \underbrace{e^*}_{\text{Effort concealment}}$$

$$\textit{Disclosure}: \mathbb{E}[e] = \sum_{(m-1)=0}^{n-1} \underbrace{\frac{(n-1)!}{(m-1)!(n-m)!} q^{m-1}(1-q)^{n-m}}_{\text{Probability of group size } m} \underbrace{\frac{(m-1)}{m^2} W}_{\text{Effort disclosure}} = \underbrace{e^*}_{\text{Effort concealment}}$$

Under concealment, conditional on one player being active, the expected effort is simply the equilibrium effort under concealment, because contestants exert the same effort for each realized group size. Under concealment, the weighted sum over the equilibrium choice for the specific realized group size is taken. This is exactly, how the equilibrium effort decision under concealment is computed (see Equation A.7). Hence, the two expected efforts are equivalent.

As the equilibrium sabotage under concealment cannot be solved analytically, I compare the numerical solutions under both policies. Figure A.1 depicts the expected sabotage, conditional on at least one player being active, under both policies and shows that there are only very small and negligible differences, if any. Therefore, I show that there are no substantial differences between average sabotage levels under concealment and disclosure.

Figure A.1: Equilibrium sabotage comparison between disclosure policy



Note: The figure depicts equilibrium expected sabotage levels under concealment compared to disclosure for low and high enter probabilities. The x-axes show the group size of all potential contestants (active and non-active). The y-axes show the average sabotage levels. Red lines indicate concealment and blue lines disclosure.

A.1.5 Expected Individual Payoff Simulation

In this section, I show that the individual expected payoff is not substantially different between disclosure policies. For this, I calculate the expected payoff conditional on participation. Because of symmetry, all players employ the same effort and sabotage. Therefore, performances are identical and all active players' winning probabilities are reduced to: $p_{win} = \frac{1}{m}$, with m being the realized number of active players. Under disclosure, conditional on participating, the expected payoff for player i is as follows:

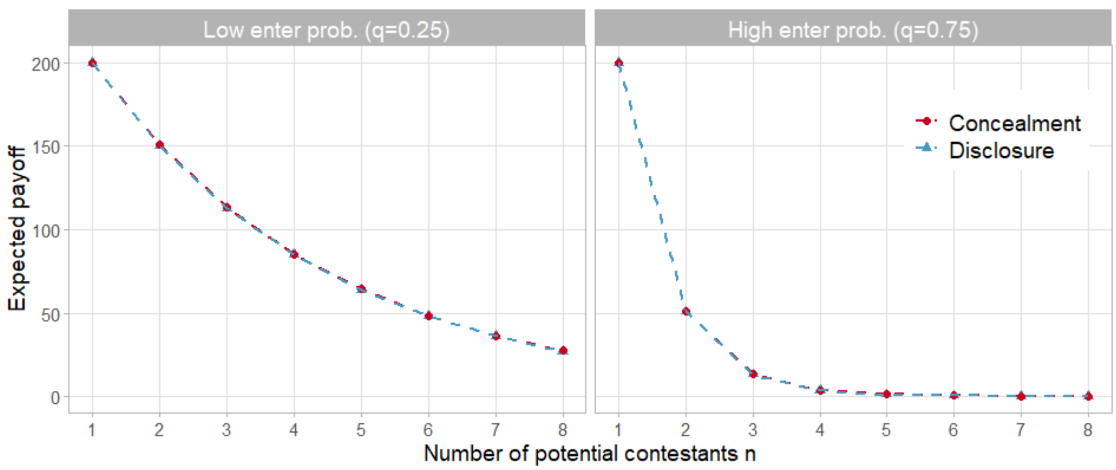
$$E[\pi_i]_{Disclosure} = \sum_{(m-1)=0}^{n-1} \left(\underbrace{\frac{(n-1)!}{(m-1)!(n-m)!} q^{m-1}(1-q)^{n-m}}_{\text{probability for } m-1 \text{ other active players}} \times \underbrace{\frac{1}{m}}_{\text{probability to win}} \times W - \underbrace{e^*(m) - s^*(m)}_{\text{costs}} \right) \tag{A.9}$$

Conditional on concealment, the expected payoff for a participant, conditional on participating is as follows:

$$E[\pi_i]_{Conceal} = \sum_{(m-1)=0}^{n-1} \Bigg(\underbrace{\frac{(n-1)!}{(m-1)!(n-m)!} q^{m-1}(1-q)^{n-m}}_{\text{probability for } m-1 \text{ other active players}} \times \underbrace{\frac{1}{m}}_{\text{probability to win}} \times W \Bigg) - \underbrace{e^*(q,n) - s^*(q,n)}_{\text{costs}} \tag{A.10}$$

Figure A.2 shows the numerical solution and indicates that there is no substantial difference between the individual ex-ante expected payoffs between disclosing and concealing the number of participants. It further shows that expected payoffs decrease both in the number of potential contestants and in their enter probabilities. This is because players win with certainty if they are the only contestants, and the probability of being alone in the contest decreases with an increasing number of potential contestants and enter probabilities.

Figure A.2: Equilibrium payoff comparison between disclosure policy



Note: The figure depicts equilibrium expected payoffs under concealment compared to disclosure for low (left graph) and high (right graph) enter probabilities. The x-axes show the group size of all potential contestants (active and non-active). The y-axes depict the expected payoffs. Red lines indicate concealment and blue lines disclosure of the number of active contestants.

A.1.6 Robustness Effectiveness of Sabotage

This section provides a robustness check for the theoretical results, with a slightly different performance function that allows for a different effectiveness in the received sabotage. Specifically, I use the following performance function:

$$y_i = \frac{e_i}{(1+(m-1)s)^t}$$

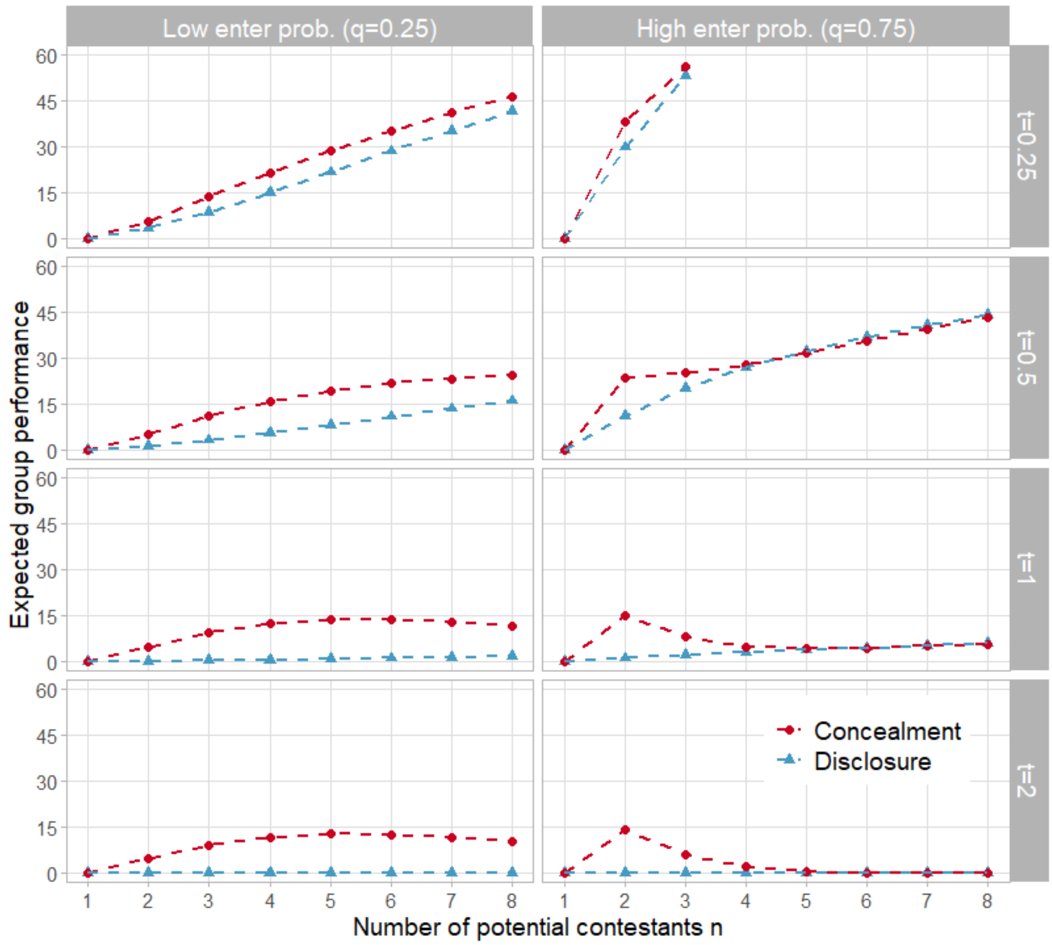
I show that the difference in the expected group performance still holds for different parameters of t. As $\lim_{t\to 0}$, however, the difference disappears. Yet, this is not surprising, because $\lim_{t\to 0}$ means that sabotage has no influence on the performance overall and thus cannot induce differences between the disclosure policies. When contestants know the number of active contestants m, The effort and sabotage levels in the symmetric equilibrium look as follows:

$$e^* = \frac{(m-1)}{m^2}W$$

$$s^* = \begin{cases} t\frac{1}{m^2}W - \frac{1}{m-1} & \text{if } W \geq \frac{m^2}{t(m-1)} \text{ and } m \geq 2 \\ 0 & \text{else} \end{cases}$$

Note that only the equilibrium sabotage levels are affected by the effectiveness of sabotage parameter t. Specifically, the less effective (higher t) sabotage, the lower are equilibrium sabotage levels. As sabotage levels are (almost) the weighted sum of the group size disclosure levels, a lower t also leads to lower sabotage levels under concealment. Figure A.3 shows the numerical solutions of the expected group performances for different levels of t, i.e., $t \in \{0.25, 0.5, 1, 2\}$. It shows that even for a small effectiveness of sabotage, which also induces lower sabotage levels overall, the difference between concealment and disclosure is still pronounced. However, this difference becomes smaller, the smaller t, with no differences as $\lim_{t\to 0}$. Nonetheless, for not too low values of t, the differences are pronounced between the disclosure policies. Therefore, this analysis shows that the differences in the group performance between the disclosure policies remain robust, even if the performance function does not carry as pronounced decreasing marginal return in the received sabotage.

Figure A.3: Equilibrium expected group performance per effectiveness of sabotage



Note: the figure depicts equilibrium expected group performance (sum of individual performances) conditional on the disclosure policy for low (left panel) and high (right panel) enter probabilities. The four lines vary the effectiveness of sabotage parameter $t \in \{0.25, 0.5, 1, 2\}$. The prize W is set to 200.

A.1.7 Preference for Donations

Suppose that agents have a preference for donations D. Specifically, suppose that agents' utility from donations is described by $U_{donations} = \alpha D$ and for simplicity that the utility gains from own payoff and the donations are additive, such that $U_i = \pi_i + \alpha D$ with $\alpha \in [0, 1)$. The equilibrium levels under group size disclosure and group size uncertainty are only marginally influenced by the preference for donations parameter α. As a consequence, the comparative statics remain the same. Furthermore, the expected group performances and thus the difference between the expected performances between the disclosure policies are also only marginally affected by this preference for donations parameter. In conclusion, the main theoretical comparative statics are robust to the inclusion of preferences for donation.

Group Size Disclosure

Conditional on the realized group size m with the set of active player M, the overall expected utility under group size disclosure is then described by:

$$\mathbb{E}[U_i] = \frac{y_i}{\sum_{j\in M} y_j} W + \alpha(\sum_{j\in M} y_j + 10) - e_i - s_i$$

with $y_i = \frac{e_i}{1+\sum_{j\neq i} s_j}$ and $\frac{y_i}{\sum_{j\in M} y_j} = \frac{1}{n}$, if $y_i = 0 \ \forall i \in M$. Suppose, agents simultaneously maximize their expected utility by choosing e_i and s_i. Then the equilibrium effort and sabotage levels can be described by the following two equations:

$$e^* = \frac{-\sqrt{(m-1)^2 W^2 - 4(m-1)\alpha m^2 \left(m-\frac{1}{2}\right) W + \alpha^2 m^4} + (2W-\alpha) m^2 - 3Wm + W}{2m^2 (m-1)}$$

$$s^* = \frac{\sqrt{2\sqrt{(m-1)^2 W^2 - 4a(m-1)\left(m-\frac{1}{2}\right) m^2 W + a^2 m^4}\, a m^2 + (m-1)^2 W^2 - 4a(m-1)\left(m-\frac{1}{2}\right) m^2 W + 2a^2 m^4}}{2m^2 (m-1)} + \frac{Wm - 2m^2 - W}{2m^2 (m-1)}$$

Proof. Suppose that agents simultaneously maximize their expected payoff:

$$\mathbb{E}[\pi_i] = \frac{y_i}{\sum_{j\in M} y_j} W + \alpha(\sum_{j\in M} y_j + 10) - e_i - s_i$$

with $y_i = \frac{e_i}{1+\sum_{j\neq i} s_j}$ and $\frac{y_i}{\sum_{j\in M} y_j} = \frac{1}{m}$, if $y_i = 0 \ \forall i \in M$. First, I take the first order condition of the expected profit function with respect to e_i:

$$\frac{\partial \pi_i}{\partial e_i} = \frac{\left(\frac{1}{1+\sum_{j\neq i} s_j}\right) \sum_{j=1}^{m} \left(\frac{e_j}{1+\sum_{l\neq j} s_l}\right) - \left(\frac{e_i}{1+\sum_{j\neq i} s_j}\right)\left(\frac{1}{1+\sum_{j\neq i} s_j}\right)}{\left(\sum_{j=1}^{m} \frac{e_j}{1+\sum_{l\neq j} s_l}\right)^2} W + \alpha \frac{1}{1+\sum_{j\neq i} s_j} - 1 = 0$$

Applying symmetry yields:

$$\frac{(m-1)}{m^2} \frac{W}{e} + \alpha \frac{1}{1+(m-1)s} = 1 \tag{A.11}$$

Next, suppose without loss of generality that player i is the m-th player. I then take the first order condition with respect to s_i:

$$\frac{\partial \pi_i}{\partial s_i} = \frac{(\frac{e_i}{(1+\sum_{j\neq i} s_j)})(\frac{e_1}{(1+\sum_{l\neq 1} s_l)^2} + ... + \frac{e_{m-1}}{(1+\sum_{l\neq m-1} s_l)^2})}{(\sum_{j=1}^{m} \frac{e_j}{1+\sum_{l\neq j} s_l})^2} W$$
$$+\alpha(-\frac{e_1}{(1+\sum_{l\neq 1} s_l)^2} - ... - \frac{e_{m-1}}{(1+\sum_{l\neq m-1} s_l)^2}) - 1 = 0$$

Symmetry yields:

$$\frac{(m-1)}{m^2}\frac{W}{(1+(m-1)s)} - \alpha(m-1)\frac{e}{(1+(m-1)s)^2} = 1$$
$$\implies s_{1,2} = \frac{(m-1)W - 2m^2 \pm \sqrt{(m-1)^2W^2 - 4\alpha m^4(m-1)e}}{2m^2(m-1)}$$

As this yields two solutions, I check which of the two is admissible. For this I plug in $\alpha = 0$ to see whether the expression collapses to the solution without preferences for donations. This is only true for $s_1 = \frac{(m-1)W-2m^2+\sqrt{(m-1)^2W^2-4\alpha m^4(m-1)e}}{2m^2(m-1)}$. I now take s_1 and plug it into Equation A.11:

$$\frac{(m-1)}{m^2}\frac{W}{e} + \alpha\frac{1}{1+(m-1)\frac{(m-1)W-2m^2+\sqrt{(m-1)^2W^2-4\alpha m^4(m-1)e}}{2m^2(m-1)}} = 1$$
$$\implies e_{1,2} = \frac{\pm\sqrt{(m-1)^2W^2 - 4(m-1)\alpha m^2\left(m-\frac{1}{2}\right)W + \alpha^2m^4} + (2W-\alpha)m^2 - 3Wm + W}{2m^2(m-1)}$$

Similarly, to determine, which of the two solutions for e is admissible, I plug in $\alpha = 0$ and see, if the solution relaxes to $e = \frac{(m-1)}{m^2}W$, which is the case without any preferences for donations. This is only the case for:

$$e^* = \frac{-\sqrt{(m-1)^2W^2 - 4(m-1)\alpha m^2\left(m-\frac{1}{2}\right)W + \alpha^2m^4} + (2W-\alpha)m^2 - 3Wm + W}{2m^2(m-1)}$$

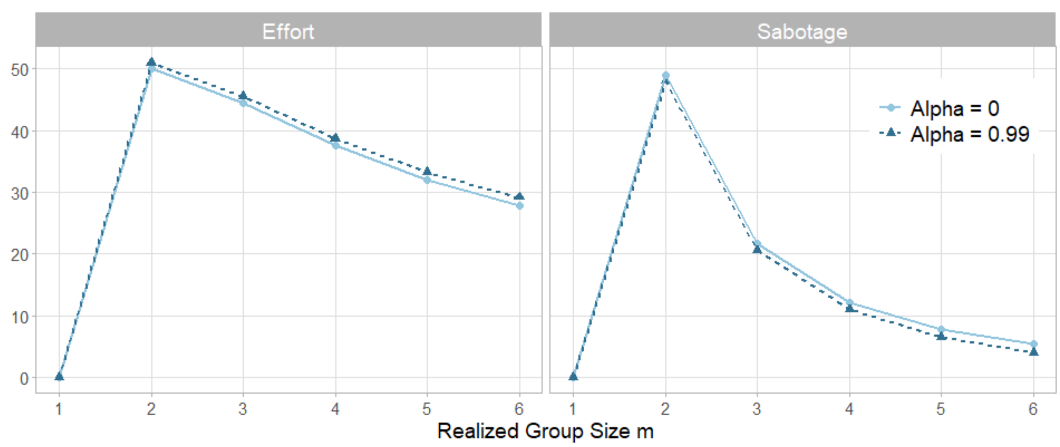
Plugging this into s_1 yields:

$$s^* = \frac{\sqrt{2\sqrt{(m-1)^2 W^2 - 4a(m-1)\left(m-\frac{1}{2}\right) m^2 W + a^2 m^4}\, a\, m^2 + (m-1)^2 W^2 - 4a(m-1)\left(m-\frac{1}{2}\right) m^2 W + 2a^2 m^4}}{2m^2(m-1)} + \frac{Wm - 2m^2 - W}{2m^2(m-1)}$$

which relaxes to $s = \frac{1}{m^2}W - \frac{1}{m-1}$ for $\alpha = 0$ and hence is admissible. □

Figure A.4 illustrates the equilibrium effort and sabotage levels for a preference for donations parameters of $\alpha = 0$ and $\alpha = 0.99$. It shows that a preference for donations only marginally changes the equilibrium choices. Specifically, effort levels are slightly higher and sabotage levels slightly lower. More importantly, a preference for donations does not change the comparative statics of the realized group size. This is because the marginal benefit of increasing the winning probability of the prize W through higher effort is much greater than the marginal benefit of more donations through lower sabotage.

Figure A.4: Equilibrium effort and sabotage under group size disclosure per preference for donation



Note: The figure depicts equilibrium effort and sabotage levels conditional on the known realized group size m and the preference for donation parameter α. Dashed dark blue lines indicate a high preference for donations and light blue lines no preference for donations.

Group Size Concealment

The expected utility with a preference for donations under group size concealment is as follows:

$$\mathbb{E}[\pi_i] = \sum_{M_i \in \mathcal{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|}\left[\frac{y_i}{\sum_{j\in M} y_j}W + \alpha\left(\sum_{j\in M} y_j + 10\right)\right] - e_i - s_i$$

with $y_i = \frac{e_i}{1+\sum_{j\neq i} s_j}$ and $\frac{y_i}{\sum_{j\in M} y_j} = \frac{1}{m}$, if $y_i = 0\ \forall i \in M$. The first order condition of the expected profit function with respect to e_i is:

$$\frac{\partial \pi_i}{\partial e_i} = \sum_{M_i \in \mathcal{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|}[\frac{(\frac{1}{1+\sum_{j\neq i} s_j)})\sum_{j=1}^{m}(\frac{e_j}{1+\sum_{l\neq j} s_l)}) - (\frac{e_i}{1+\sum_{j\neq i} s_j})(\frac{1}{1+\sum_{j\neq i} s_j})}{(\sum_{j=1}^{m}\frac{e_j}{1+\sum_{l\neq j} s_l})^2}W$$

$$+\alpha\frac{1}{1+\sum_{j\neq i} s_j}] - 1 = 0$$

I now apply symmetry. With homogenous contestants, the sum over all possible sets of all other active contestants. relaxes to all possible number of others. For readability, I define $B^{n-1}_{m-1} = \sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-1-(m-1))!} q^{(m-1)}(1-q)^{n-1-(m-1)}$:

$$B^{n-1}_{m-1}\left[\frac{(m-1)}{m^2}\frac{W}{e} + \alpha\frac{1}{1+(m-1)s}\right] = 1$$

$$\iff e = \frac{B^{n-1}_{m-1}\frac{m-1}{m^2}W}{1 - B^{n-1}_{m-1}\alpha\frac{1}{1+(m-1)s}} \tag{A.12}$$

Next, suppose without loss of generality that player i is the m-th player. The first order condition with respect to s_i is:

$$\frac{\partial \pi_i}{\partial s_i} = \sum_{M_i \in \mathcal{P}^{N_i}} q^{|M_i|}(1-q)^{|N_i/M_i|}[\frac{(\frac{e_i}{(1+\sum_{j\neq i} s_j)})(\frac{e_1}{(1+\sum_{l\neq 1} s_l)^2} + \ldots + \frac{e_{m-1}}{(1+\sum_{l\neq m-1} s_l)^2})}{(\sum_{j=1}^{m}\frac{e_j}{1+\sum_{l\neq j} s_l})^2}W$$

$$+\alpha(-\frac{e_1}{(1+\sum_{l\neq 1} s_l)^2} - \ldots - \frac{e_{m-1}}{(1+\sum_{l\neq m-1} s_l)^2})] - 1 = 0$$

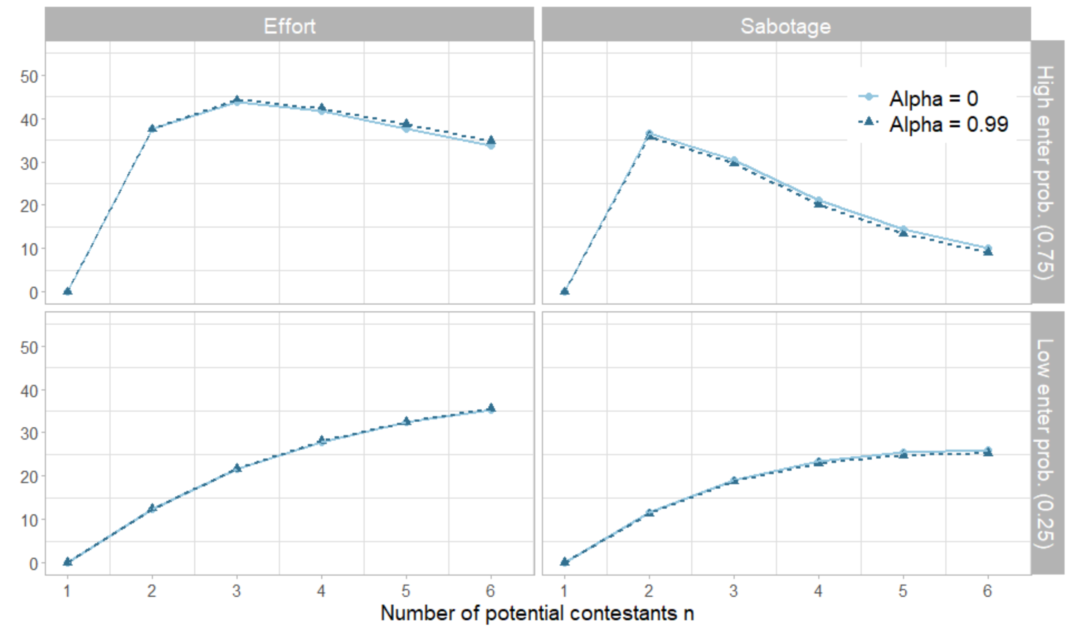
Applying symmetry, and again with $B^{n-1}_{m-1} = \sum_{(m-1)=0}^{n-1} \frac{(n-1)!}{(m-1)!(n-1-(m-1))!} q^{(m-1)}(1-q)^{n-1-(m-1)}$, it becomes:

$$B^{n-1}_{m-1}\left[\frac{(m-1)}{m^2}\frac{W}{(1+(m-1)s)} - \alpha(m-1)\frac{e}{(1+(m-1)s)^2}\right] = 1 \tag{A.13}$$

Which does not yield a closed-form solution for s. Hence, I solve equations A.12 and A.13 numerically. Figure A.5 shows the numerical solution for the parameters of interest. It shows the equilibrium effort and sabotage levels for a preference for donations parameter of $\alpha = 0$

and $\alpha = 0.99$. It reveals only marginal differences in the effort and sabotage levels between these parameters.

Figure A.5: Equilibrium effort and sabotage under group size uncertainty per preference for donation

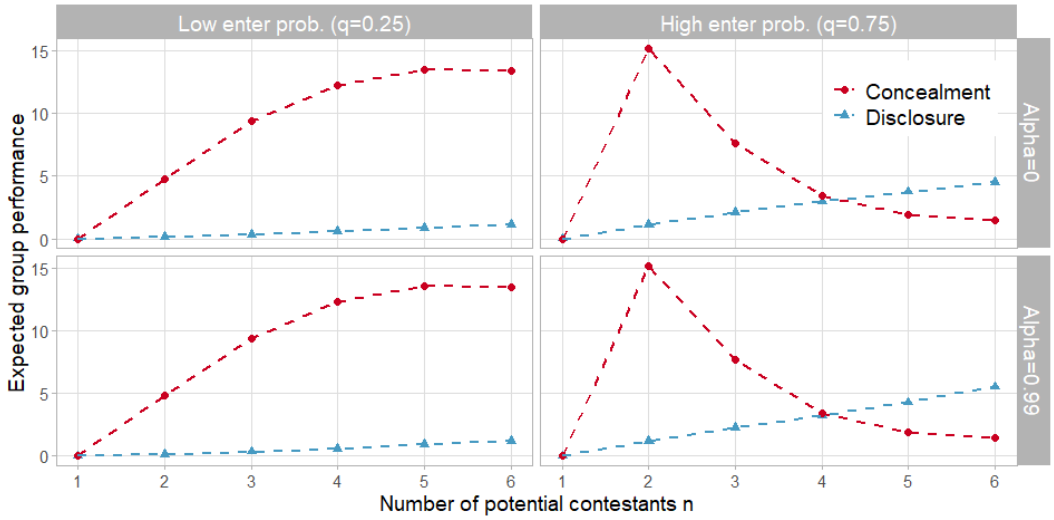


Note: The panels depicts equilibrium effort and sabotage levels under group size uncertainty conditional on the preference for donation parameter α. Dashed dark blue lines indicate a high preference for donations and light blue lines no preference for donations.

Comparison Disclosure Policies

Figure A.6 depicts the difference in expected group performances between concealment and disclosure for a preference for donation parameter of $\alpha = 0$ (upper row) and $\alpha = 0.99$. The difference between the disclosure policies is almost the same between the two parameters. Consequently, a preference for donation parameter does also not change this comparative static.

Figure A.6: Equilibrium group performance comparison between disclosure policy per preference for donation

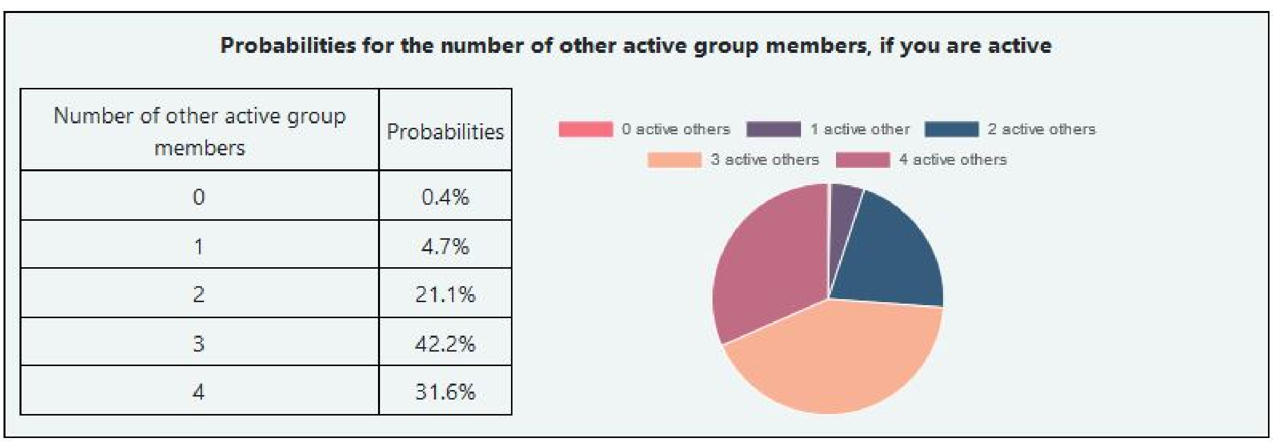


Note: The panels depict the expected group performance under concealment (red) and disclosure (blue) for low enter probabilities (left) and high enter probabilities (right) for either $\alpha = 0$ (upper row) or $\alpha = 0.99$ (lower row).

A.2 Experimental Design Appendix

Figure A.7 shows the communicated group size probabilities in Part B of the experiment for Treatment *5H*. For all other treatments, this looked the same but only with the respective probabilities and possible number of active group members (only 0, 1, 2 for Treatment *3L* and Treatment *3H*).

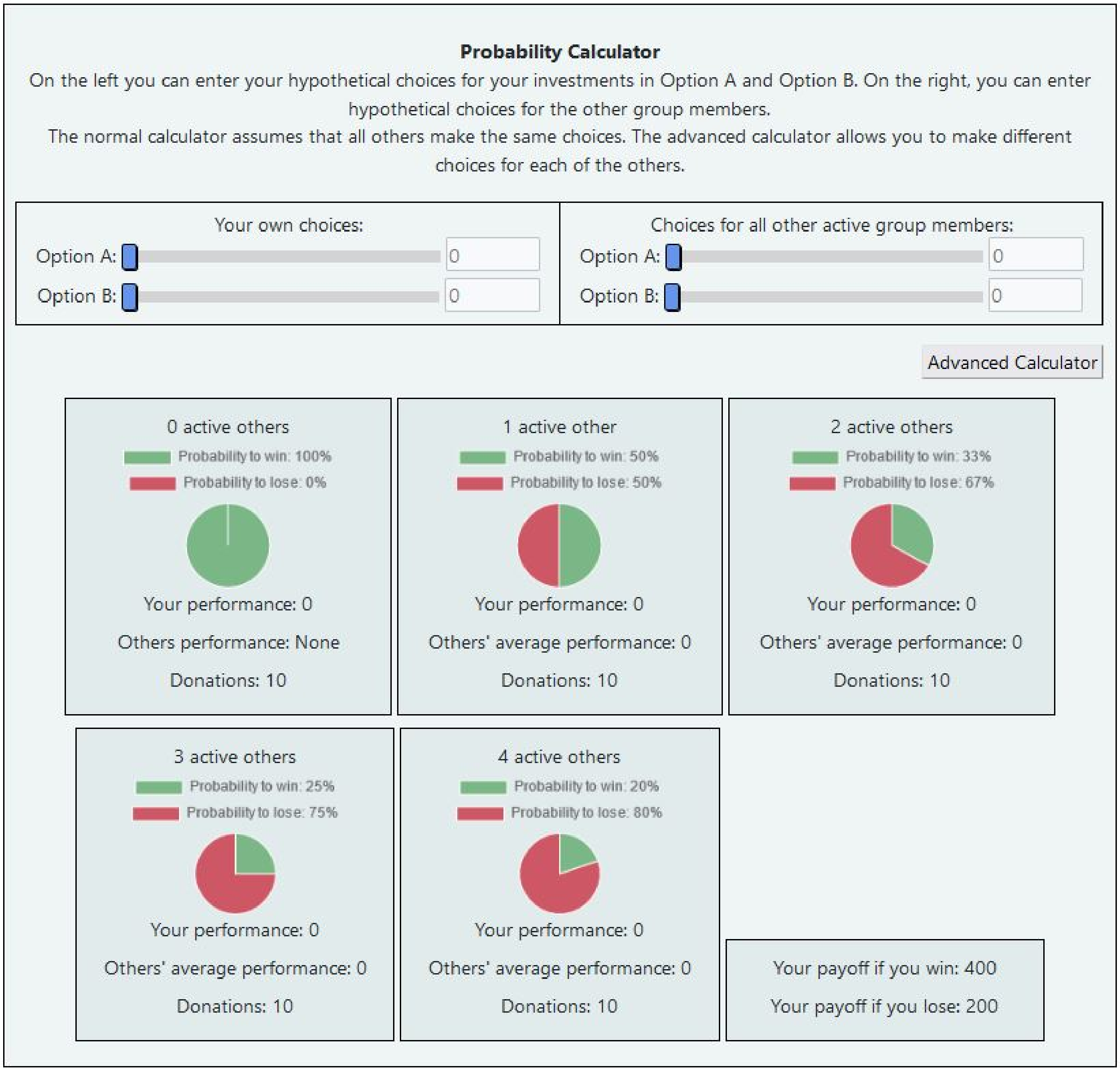
Figure A.7: Communicated group size probabilities in Part B (for Treatment *5H*)



Number of other active group members	Probabilities
0	0.4%
1	4.7%
2	21.1%
3	42.2%
4	31.6%

Figure A.8 shows the probability calculator that subjects had access to at any time. By clicking on the button 'advanced calculator', they could enter Option-A and Option-B choices for each of their potential competitors individually.

Figure A.8: Probability calculator



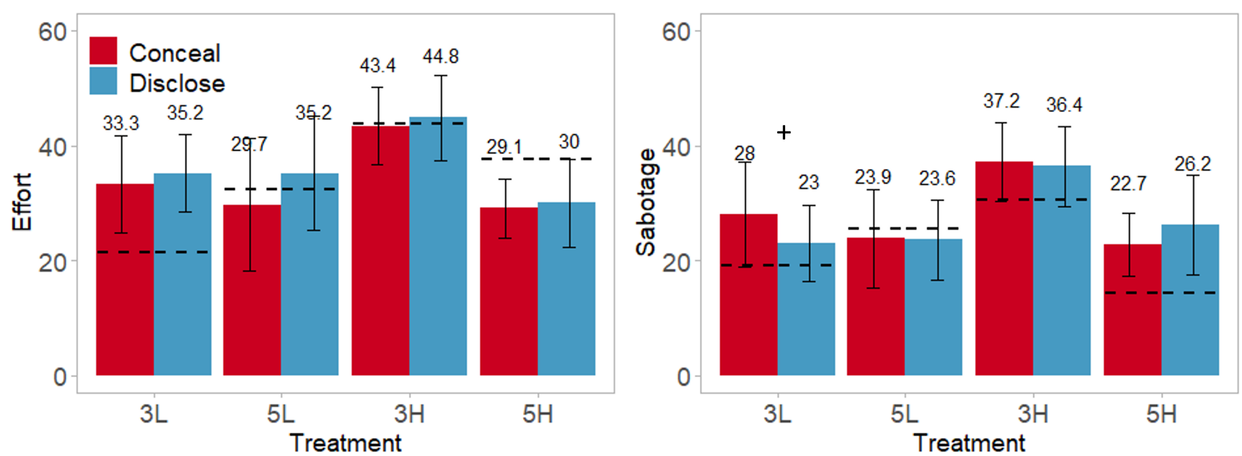
A.3 Results Appendix

A.3.1 Disclosure Policy

Effort, Sabotage, and Expected Payoff per Treatment

Figure A.9 shows effort and sabotage differences across the policies conditional on the treatments. It shows that there are no significant differences in effort and sabotage levels between the disclosure policies for any of the treatments, except in Treatment *3L*, where concealment increases sabotage levels ($p < 0.1$). The robustness checks confirm these results (see Appendix A.3.1), and find a significant ($p < 0.05$) increase in sabotage under concealment for Treatment *3L* treatment.[1]

Figure A.9: Results effort, sabotage, and payoff per treatments between disclosure policy

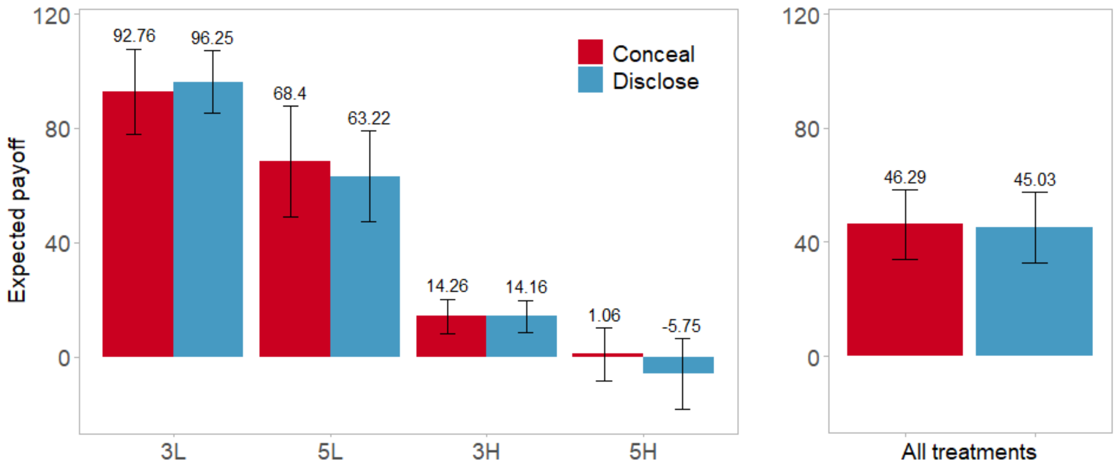


Note: The bar charts depict the average effort (left panel) and sabotage (right panel) conditional on the disclosure policy and the treatments. Black dashed lines show the Nash equilibrium predictions. The error bars show 95% confidence intervals. Significance levels: + $p < 0.10$

Figure A.10 compares the expected payoff between the disclosure policies for each treatment (left panel) and pooled over all treatments (right panel). It shows that there are no significant differences between the disclosure policies for all treatments. The robustness checks also do not find any significant differences (see Appendix A.3.1).

[1] Specifically, I find a significant ($p < 0.05$) increase under concealment for Treatment *3L*, when studying only the rounds around the policy changes and in the regression analysis. Apart from that, there is no significant difference in effort and sabotage levels between the disclosure policies in the robustness checks.

Figure A.10: Results expected payoffs between disclosure policy per treatment



Note: The figure depicts the individual expected payoffs based on the subjects' choices conditional on the treatments (left panel) or pooled over all treatments (right panel). Error bars show 95% confidence intervals.

Time Trends

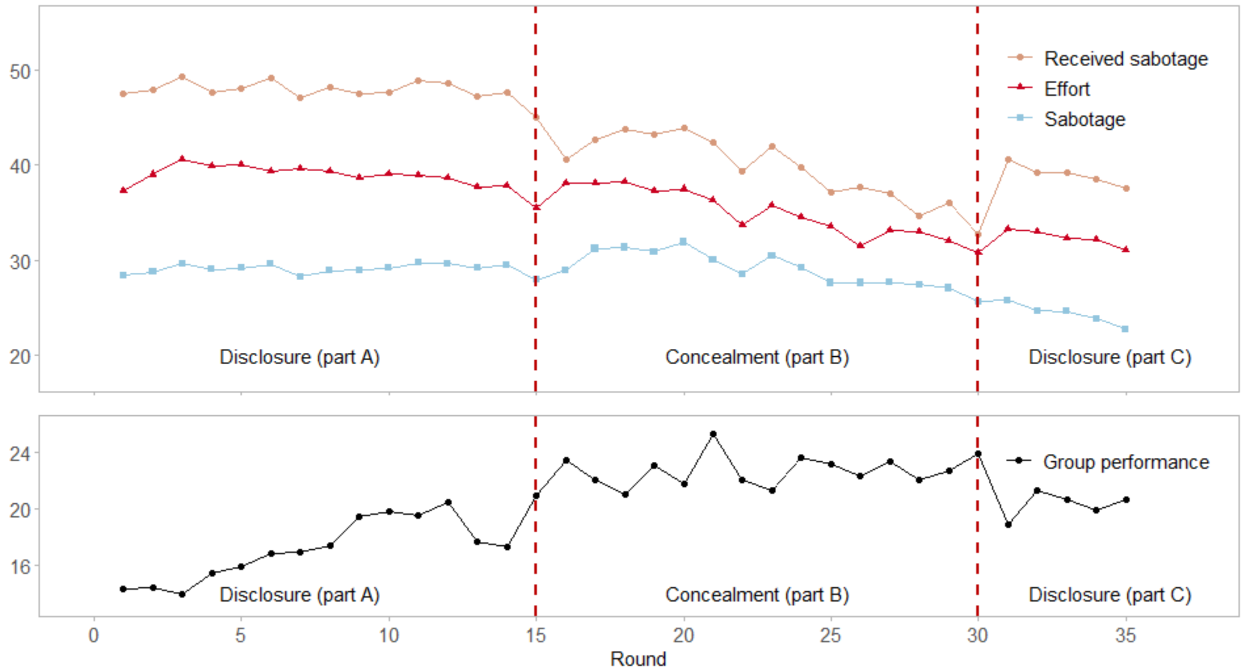
Figure A.11 depicts the time trends of the expected received sabotage, the expected effort and sabotage levels, as well as the expected group performance over all rounds and pooled over all treatments. The upper panel shows that there is a slight decrease in the expected choices over time, whereas the lower panel shows that there is a slight increase in the expected group performance over time. Therefore, I conduct the two robustness checks, where I first focus only on the rounds around the disclosure policy changes and on regression analyses that control for the time trend. The robustness checks support the results of the main section.

Robustness Check Effort, Sabotage and Expected Payoff

Subset of Rounds around Policy Change

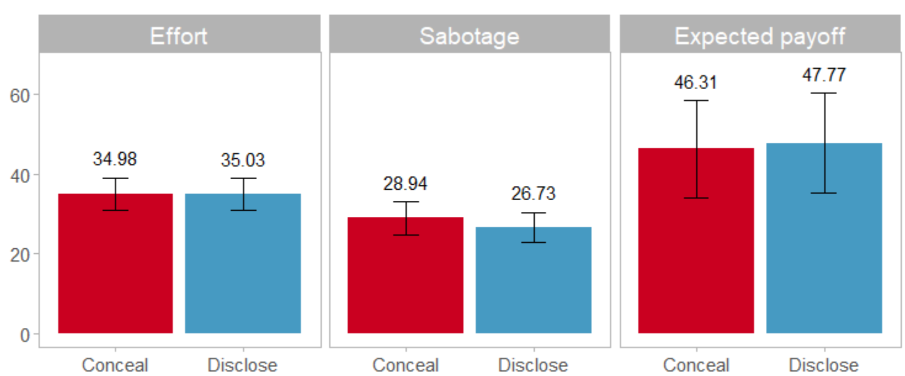
In this section, I focus on the rounds that are in the neighborhood of the disclosure policies, i.e. rounds 11-20 and 21-30. The following figure shows the pooled averages only for those rounds. Figure A.12 shows the pooled averages over all treatments. It shows no significant differences in average expected effort, expected sabotage levels, and expected payoffs between the disclosure policies. Figure A.13 shows average expected effort and sabotage levels conditional on the disclosure policy and each treatment. It shows no significant differences in levels between the policies, except for Treatment *3L*, where concealment leads to significantly ($p < 0.05$) higher sabotage levels. Similarly, Figure A.14 shows the expected payoff conditional on the disclosure policy for each treatment individually. It, too, shows no significant difference between the disclosure policies.

Figure A.11: Overview time trends



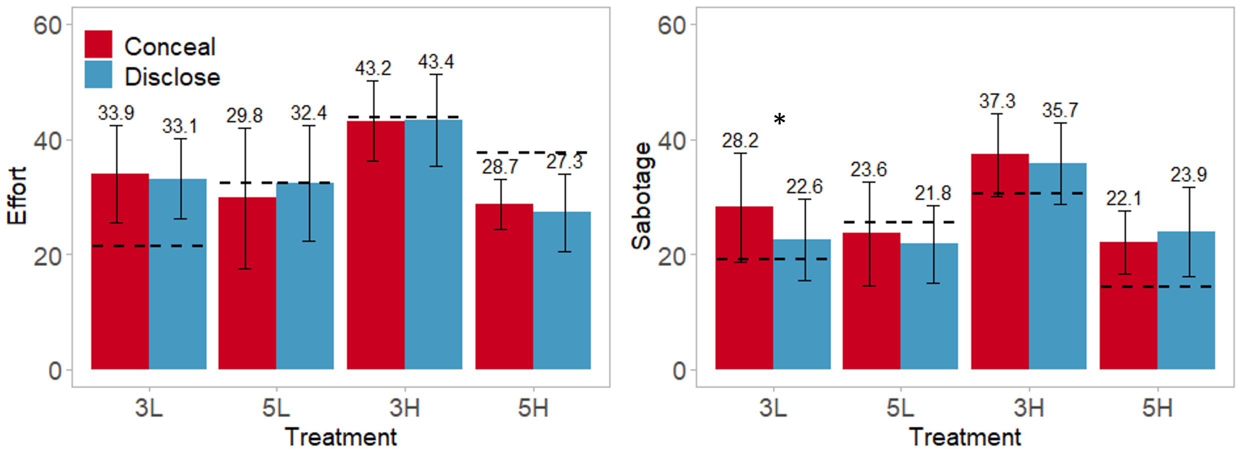
Note: The two panels show the average expected received sabotage, the average expected effort and sabotage levels, and the average expected group performance, based on the subjects' choices over all rounds.

Figure A.12: Robustness results average expected effort, sabotage, and payoff per disclosure policy



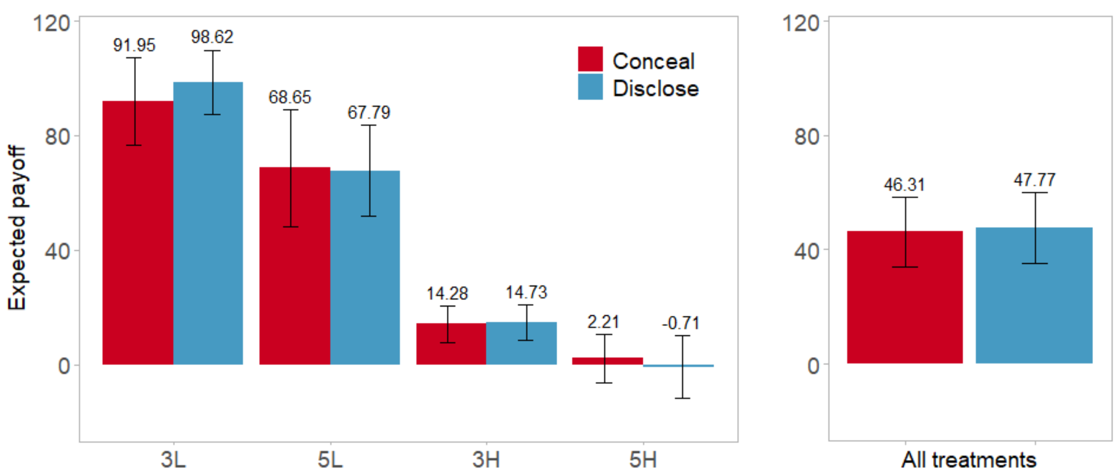
Note: The figure shows average effort, sabotage, and expected payoffs conditional on the disclosure policy, pooled over all treatments. The figure is based solely on rounds 11-20 and 21-30 (around policy changes). Error bars show 95% confidence intervals. Significance levels: + $p < 0.10$

Figure A.13: Robustness results effort, sabotage, and payoffs per treatment between disclosure policy



*Note: The bar charts show the average effort (left panel) and sabotage (right panel) conditional on the disclosure policy and the treatments for rounds 11-20 and 21-30 (around policy changes). The error bars show 95% confidence intervals. Significance levels: * $p < 0.05$*

Figure A.14: Robustness results expected payoffs between discloure policy per treatment



Note: The figure depicts the average expected individual payoffs based on the subjects' choices conditional on the treatments (left panel) or pooled over all treatments (right panel) for the rounds 11-20 and 26-30. Error bars show 95% confidence intervals.

Regression Analyses

As an additional robustness check, I run linear regression models, which control for the time trend. I cluster standard errors at the group level. I include the pre-registered controls[2] and additionally include the treatments as controls. Table A.1 shows no significant difference in effort levels and expected payoffs between the disclosure policies. It further shows that concealment significantly ($p < 0.05$) increases sabotage levels.

Next, I run the same regressions, but conditional on the specific treatments. Table A.2 shows linear regressions of the expected average effort and sabotage levels on concealment but for each treatment separately. It confirms the significant ($p < 0.05$) increase in sabotage levels under concealment for the treatment $3L$ and additionally shows a significant ($p < 0.05$) increase in effort under concealment in Treatment *5H*. Finally, I run the same regressions but for expected payoffs. Again, it does not show any significant differences between the disclosure policies for any of the treatments (Table A.3).

[2] The controls are: being active in the round before, having won in the round before, average sabotage and effort levels of other participants in the round before, round, determined group size in the round before, how often won in the rounds before, SVO, spite, risk, loss and ambiguity aversion, age, gender, highest degree, the field of study, the degree of concentration and understanding.

Table A.1: Linear regression expected effort, sabotage, and payoff on concealment and controls

	Dependent Variable:		
	Effort	*Sabotage*	*Expected Payoff*
	(1)	(2)	(3)
Concealment	0.43	2.20*	−0.18
	(1.09)	(1.11)	(1.50)
Round	−0.59***	−0.40***	0.65***
	(0.10)	(0.08)	(0.11)
Risk Aversion	−1.72	−1.64	1.18
	(1.51)	(1.34)	(1.82)
Loss Aversion	−0.59	−1.03	−1.54
	(0.86)	(0.88)	(1.02)
Ambiguity Aversion	−3.09*	−2.04+	3.34**
	(1.35)	(1.08)	(1.26)
SVO	0.19	0.11	−0.19
	(0.18)	(0.17)	(0.17)
Spite	10.43	18.35*	0.63
	(8.75)	(8.89)	(9.63)
Female	6.99+	5.87+	−6.03+
	(3.94)	(3.39)	(3.53)
Age	0.18	−0.58	−0.42
	(0.63)	(0.48)	(0.63)
Constant	46.39*	64.84***	39.78*
	(20.75)	(18.75)	(20.10)
Treatment Dummies	✓	✓	✓
Other Controls	✓	✓	✓
Observations	6,630	6,630	6,630
# Clusters	52	52	52
R^2	0.16	0.15	0.69

Note: SE clustered at group level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table A.2: Linear regression expected effort and sabotage on disclosure and controls for different treatments

	Dependent Variable:							
	Effort	*Sabotage*	*Effort*	*Sabotage*	*Effort*	*Sabotage*	*Effort*	*Sabotage*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Concealment	0.58	4.98*	−0.89	3.18	0.62	1.91	3.99*	1.03
	(2.61)	(2.08)	(2.82)	(3.34)	(1.76)	(1.44)	(1.84)	(2.10)
Round	−0.28	−0.01	−0.82**	−0.64*	−0.34*	−0.34**	−1.01***	−0.76***
	(0.19)	(0.09)	(0.28)	(0.28)	(0.16)	(0.12)	(0.22)	(0.18)
Risk Aversion	−8.53**	−8.90***	−2.71*	−1.78	−5.58*	−6.75***	0.93	1.52
	(2.85)	(1.85)	(1.16)	(1.29)	(2.52)	(1.94)	(3.31)	(2.57)
Loss Aversion	2.83	2.47	−0.12	0.09	−4.30**	−3.11*	−2.06	−3.51**
	(2.31)	(2.02)	(1.47)	(1.49)	(1.58)	(1.56)	(1.41)	(1.27)
Ambiguity Aversion	−2.90	−2.99**	−2.68	−2.33	−3.00	−1.02	−0.24	0.50
	(1.89)	(0.99)	(2.19)	(2.01)	(3.16)	(3.35)	(1.57)	(1.13)
SVO	−0.05	0.09	−0.30	−0.52+	0.50	−0.04	−0.14	0.11
	(0.38)	(0.20)	(0.35)	(0.30)	(0.33)	(0.29)	(0.45)	(0.44)
Spite	11.10	9.64	47.84+	23.49	35.91+	35.65**	−3.31	31.31
	(12.18)	(10.22)	(25.69)	(25.58)	(18.58)	(12.65)	(43.03)	(34.87)
Female	−9.91	−0.79	−5.10	0.80	4.25	−1.24	17.83***	17.81***
	(6.64)	(2.99)	(11.66)	(11.13)	(7.13)	(6.60)	(3.96)	(4.05)
Age	3.57**	1.94**	−1.20	−1.08	−1.21	−1.34	−0.46	−2.44*
	(1.21)	(0.69)	(2.23)	(2.01)	(1.54)	(1.23)	(1.58)	(1.01)
Constant	−4.72	48.41**	80.16	83.76	124.48*	177.78***	72.55	86.67*
	(41.64)	(17.68)	(73.26)	(65.27)	(53.93)	(43.48)	(45.57)	(37.52)
Treatments	3L	3L	5L	5L	3H	3H	5H	5H
Other Controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	1,632	1,632	1,700	1,700	1,632	1,632	1,666	1,666
# Clusters	16	16	10	10	16	16	10	10
R^2	0.48	0.41	0.23	0.24	0.30	0.30	0.26	0.24

Note: SE clustered at group level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table A.3: Linear regression expected payoff on concealment and controls for different treatments

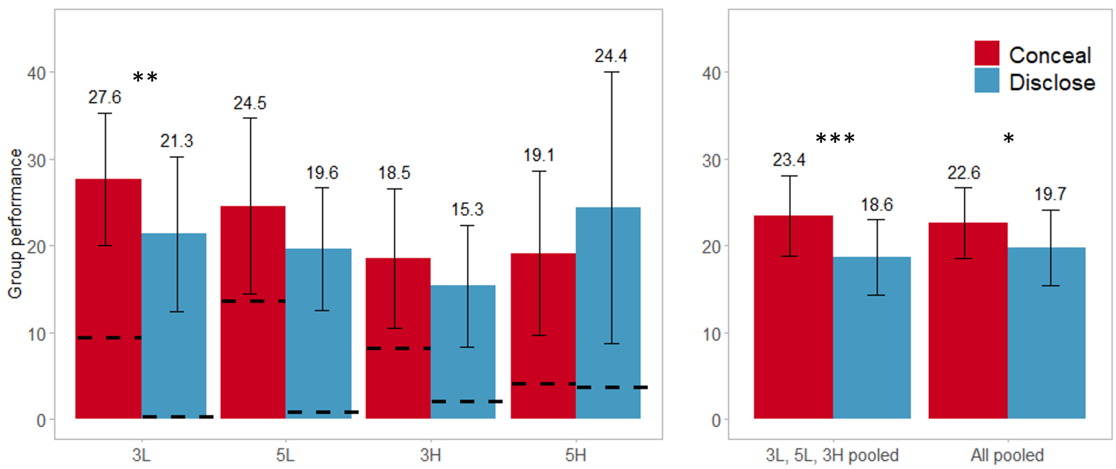
	Dependent Variable:			
	Expected Payoff			
	(1)	(2)	(3)	(4)
Concealment	−3.82	−0.68	−0.17	0.55
	(3.59)	(4.78)	(1.54)	(1.97)
Round	0.29^{+}	1.19^{**}	0.20^{*}	0.77^{***}
	(0.16)	(0.40)	(0.08)	(0.19)
Risk Aversion	12.13^{***}	0.34	2.76^{**}	0.87
	(3.18)	(1.05)	(0.86)	(2.17)
Loss Aversion	−4.30	-2.63^{+}	0.68	2.28^{+}
	(3.07)	(1.36)	(0.71)	(1.28)
Ambiguity Aversion	3.93^{*}	4.18^{+}	2.21	1.27
	(1.54)	(2.28)	(1.48)	(1.09)
SVO	0.04	0.13	−0.05	−0.15
	(0.40)	(0.35)	(0.10)	(0.27)
Spite	−15.20	11.14	14.88	−37.64
	(18.09)	(25.08)	(9.95)	(26.72)
Female	−3.02	14.94	−0.75	-13.09^{***}
	(5.10)	(9.77)	(2.17)	(2.96)
Age	-5.58^{***}	2.46^{*}	−0.34	2.49^{**}
	(1.54)	(1.11)	(0.55)	(0.76)
Constant	145.54^{***}	16.50	−2.48	-40.58^{*}
	(39.04)	(39.86)	(14.93)	(20.64)
Treatments	3L	5L	3H	5H
Other Controls	✓	✓	✓	✓
Observations	1,632	1,700	1,632	1,666
# Clusters	16	10	16	10
R^2	0.52	0.36	0.36	0.32

Note: SE clustered at group level $+\ p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Robustness Check Group Performance

Figure A.15 depicts the expected group performance conditional on the disclosure policy for each treatment separately (left panel) and over all treatments pooled, excluding and including Treatment *5H* (right panel). It shows the averages of the subsets around the policy changes, i.e., rounds 11-20 and 21-30. It confirms the results from the main section in all cases, except in the case of Treatment *3H*, where it shows no statistically significant increase under concealment. Linear regressions reveal the same effects and significance levels of concealment on group performance (see Table A.4).

Figure A.15: Robustness results group performance per treatment



*Note: The bar charts show the average group performance conditional on the disclosure policy and on the treatments. The error bars show 95% confidence intervals. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$*

Table A.4: Linear regression group performance on concealment and controls

	Dependent Variable:					
	Group Performance					
	(1)	(2)	(3)	(4)	(5)	(6)
Concealment	7.78^{**}	6.86	4.98	-5.39	3.41^{+}	6.07^{**}
	(2.43)	(4.73)	(3.48)	(3.58)	(1.89)	(2.05)
Round	-0.34^{+}	-0.73^{**}	0.25	0.73^{*}	-0.005	-0.21
	(0.18)	(0.26)	(0.21)	(0.31)	(0.15)	(0.14)
Risk Aversion	-3.92	0.43	-0.74	-2.71^{**}	-0.68	-0.24
	(2.54)	(0.58)	(0.99)	(0.92)	(0.86)	(1.04)
Loss Aversion	2.76	0.95	1.88^{*}	0.28	1.50^{***}	1.62^{**}
	(2.30)	(0.64)	(0.93)	(0.54)	(0.43)	(0.58)
Ambiguity Aversion	-1.39	-0.78	-2.45^{**}	-0.70	-1.79^{*}	-1.73^{+}
	(1.61)	(0.71)	(0.92)	(0.62)	(0.82)	(1.03)
SVO	-0.10	0.01	-0.06	-0.16	0.04	0.004
	(0.31)	(0.11)	(0.13)	(0.29)	(0.12)	(0.14)
Spite	10.36	2.76	-16.09^{***}	19.36^{*}	-0.49	-0.19
	(9.49)	(8.49)	(4.71)	(8.88)	(4.98)	(5.45)
Female	-10.68^{+}	-7.09^{+}	-4.44^{+}	1.09	-1.25	-2.18
	(6.46)	(3.74)	(2.51)	(3.42)	(2.65)	(3.06)
Age	3.33^{**}	-0.54	0.89^{+}	0.33	0.81^{*}	1.12^{*}
	(1.17)	(0.39)	(0.53)	(0.23)	(0.36)	(0.47)
Constant	-38.52	54.06^{***}	-1.53	-5.61	-14.97	-14.04
	(33.11)	(10.90)	(23.84)	(10.36)	(12.37)	(15.08)
Treatments	3L	5L	3H	5H	All	No 5H
Other Controls	✓	✓	✓	✓	✓	✓
Observations	1,632	1,700	1,632	1,666	6,630	4,964
# Clusters	16	10	16	10	52	42
R^2	0.41	0.30	0.18	0.28	0.13	0.19

Note: SE clustered at group level $+ p < 0.1$; $^{*} p < 0.05$; $^{**} p < 0.01$; $^{***} p < 0.001$

Robustness Check Implemented Choices

In this section, I replicate the main results by focusing on the choices that were implemented in each round of the experiment. Specifically, this includes only the choices for the realized group size in each round and only the participants that were chosen to become active in each round. I run several regression analyses with the pre-registered controls[3] and additionally add the implemented realized group size of the current round as a control. Table A.5 shows the regression of the implemented effort, sabotage, and resulting payoffs conditional on concealment, revealing no significant differences between the disclosure policies. Table A.6 and Table A.7 show the same regressions but for each treatment separately. It shows significantly higher sabotage levels under concealment for Treatment *3L* and significantly higher effort levels for *5L*. Apart from these significant differences, the regressions do not show any other significant differences in sabotage, effort, or payoffs between the disclosure policies. Finally, Table A.8 shows the regression of the implemented group performance on concealment and replicates the results from the main section, but other than in the main section also provides support for a significant increase in group performance for Treatment *5L*.

[3]The controls are: being active in the round before, having won in the round before, average sabotage and effort levels of other participants in the rounds before, round, determined group size in the round before, how often won in the rounds before, SVO, spite, risk, loss and ambiguity aversion, age, gender, highest degree, the field of study, the degree of concentration and understanding.

Table A.5: Linear regression effort, sabotage, and payoffs on concealment and controls for implemented choices

	Dependent Variable:		
	Effort	*Sabotage*	*Expected Payoff*
	(1)	(2)	(3)
Concealment	1.42	2.35^{+}	−1.49
	(1.39)	(1.23)	(1.89)
Round	-0.61^{***}	-0.47^{***}	0.82^{***}
	(0.11)	(0.08)	(0.12)
Risk Aversion	−1.59	−1.17	1.68
	(1.66)	(1.45)	(1.57)
Loss Aversion	−1.31	-1.76^{+}	−1.30
	(0.86)	(0.90)	(1.27)
Ambiguity Aversion	-2.40^{+}	−1.17	1.89
	(1.35)	(1.19)	(1.49)
SVO	0.18	0.11	0.07
	(0.20)	(0.19)	(0.26)
Spite	8.27	18.19^{+}	15.11
	(9.67)	(9.62)	(14.14)
Female	7.66^{*}	6.11^{+}	-10.97^{**}
	(3.48)	(3.20)	(4.22)
Age	0.16	−0.61	−0.51
	(0.64)	(0.53)	(0.83)
Constant	43.48^{+}	61.39^{**}	345.35^{***}
	(22.32)	(21.35)	(28.16)
Treatment Dummies	✓	✓	✓
Group Size Realization Dummies	✓	✓	✓
Other Controls	✓	✓	✓
Observations	3,751	3,751	3,751
# Clusters	52	52	52
R^2	0.18	0.16	0.41

Note: SE clustered at group level $+\ p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table A.6: Linear regression effort, sabotage, and payoffs on concealment and controls for implemented choices for different treatments

	Dependent Variable:							
	Effort	*Sabotage*	*Effort*	*Sabotage*	*Effort*	*Sabotage*	*Effort*	*Sabotage*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Concealment	1.73	6.81*	0.68	2.73	0.62	2.63	5.51**	2.22
	(2.83)	(2.74)	(4.05)	(3.82)	(2.62)	(1.84)	(2.02)	(2.37)
Round	−0.62**	−0.48*	−0.36	−0.53	−0.32*	−0.29+	−1.11***	−0.81***
	(0.22)	(0.19)	(0.53)	(0.37)	(0.16)	(0.15)	(0.22)	(0.17)
Risk Aversion	−8.55***	−8.07**	−2.19	−1.94	−6.03*	−7.72***	1.20	1.86
	(2.14)	(2.48)	(1.36)	(1.39)	(2.48)	(1.97)	(3.24)	(2.50)
Loss Aversion	3.77+	2.09	1.66	2.04	−4.73**	−3.46+	−1.93	−3.35**
	(2.23)	(2.84)	(1.36)	(1.44)	(1.55)	(1.78)	(1.34)	(1.14)
Ambiguity Aversion	−2.06	−1.69	−5.31**	−3.95*	−2.93	−1.16	0.17	0.91
	(1.62)	(1.35)	(1.85)	(1.66)	(3.10)	(3.32)	(1.54)	(1.03)
SVO	−0.22	−0.03	−0.36	−0.53+	0.56	−0.02	−0.27	0.04
	(0.37)	(0.25)	(0.34)	(0.31)	(0.34)	(0.31)	(0.45)	(0.44)
Spite	7.56	4.78	13.58	0.27	37.35+	36.68**	−11.52	23.63
	(12.83)	(10.67)	(22.80)	(23.10)	(19.59)	(13.11)	(40.85)	(32.63)
Female	−14.24*	−1.27	0.62	3.75	4.57	−0.65	14.67***	16.01***
	(6.55)	(3.58)	(9.87)	(8.59)	(7.54)	(6.98)	(3.27)	(3.46)
Age	4.33**	1.99*	−1.61	−1.89	−0.84	−0.82	−0.15	−2.31*
	(1.34)	(0.78)	(2.08)	(1.75)	(1.55)	(1.25)	(1.58)	(0.96)
Constant	11.55	69.00*	95.21	108.82+	118.75*	179.86***	69.56	89.37*
	(41.95)	(29.38)	(64.42)	(55.40)	(57.06)	(48.30)	(47.70)	(39.94)
Treatments	3L	3L	5L	5L	3H	3H	5H	5H
Other Controls	✓	✓	✓	✓	✓	✓	✓	✓
Observations	713	713	544	544	1,237	1,237	1,257	1,257
# Clusters	16	16	10	10	16	16	10	10
R^2	0.39	0.22	0.20	0.22	0.27	0.27	0.25	0.22

Note: SE clustered at group level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table A.7: Linear regression implemented payoff on concealment and controls

	Dependent Variable:			
	Payoff			
	(1)	(2)	(3)	(4)
Concealment	−1.02	1.28	−2.67	−3.78
	(4.26)	(8.58)	(3.16)	(2.45)
Round	1.03**	0.14	0.52*	0.91***
	(0.39)	(0.79)	(0.22)	(0.18)
Risk Aversion	9.35***	0.51	7.03*	2.47
	(2.33)	(2.36)	(2.73)	(2.74)
Loss Aversion	−5.81+	−5.91**	5.22**	−0.08
	(3.23)	(2.25)	(1.76)	(1.25)
Ambiguity Aversion	2.99+	3.90	5.24	−0.19
	(1.66)	(2.89)	(3.31)	(1.57)
SVO	1.23*	0.53	−0.14	−0.03
	(0.50)	(0.49)	(0.40)	(0.33)
Spite	12.61	46.86	14.28	−10.11
	(23.28)	(39.48)	(18.32)	(24.51)
Female	−4.93	8.41*	−3.34	−18.17***
	(9.84)	(3.95)	(7.49)	(4.62)
Age	−7.62***	2.35+	−2.55+	2.29**
	(1.92)	(1.30)	(1.49)	(0.79)
Constant	348.62***	325.98***	313.77***	324.74***
	(69.24)	(44.88)	(49.81)	(28.59)
Treatments	3L	5L	3H	5H
Other Controls	✓	✓	✓	✓
Observations	713	544	1,237	1,257
# Clusters	16	10	16	10
R^2	0.59	0.50	0.24	0.10

Note: SE clustered at group level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table A.8: Linear regression implemented group performance on concealment and controls

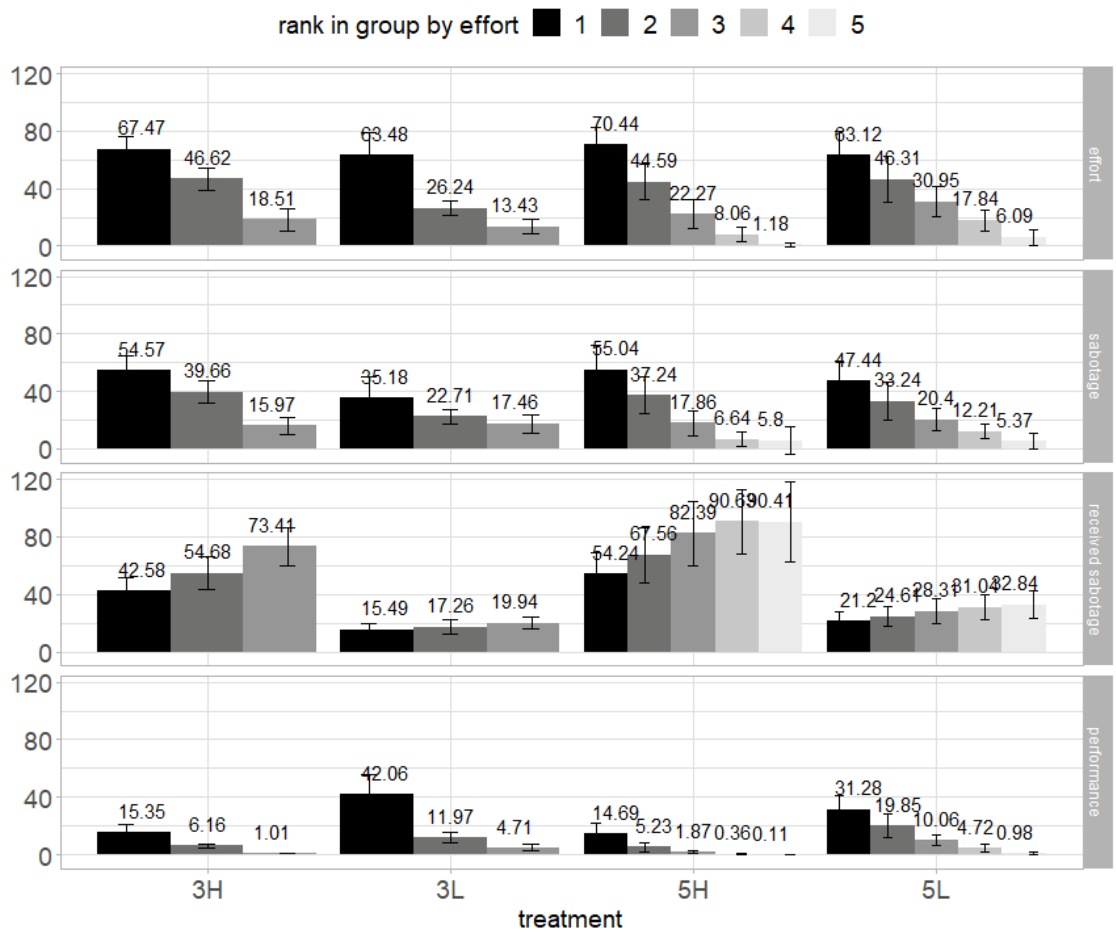
	Dependent Variable:					
	Group Performance					
	(1)	(2)	(3)	(4)	(5)	(6)
Concealment	6.34^{+}	14.84*	7.88**	−7.39^{+}	5.60*	9.51***
	(3.28)	(6.34)	(3.03)	(3.78)	(2.55)	(2.73)
Round	−0.18	−0.34	−0.03	0.59^{+}	0.05	−0.13
	(0.11)	(0.39)	(0.20)	(0.33)	(0.14)	(0.11)
Risk Aversion	−0.31	−0.33	−1.05	−3.26***	−0.73	−0.41
	(1.06)	(0.47)	(0.75)	(0.98)	(0.52)	(0.55)
Loss Aversion	1.40	0.31	2.41***	0.05	0.89**	0.99**
	(1.08)	(0.32)	(0.63)	(0.48)	(0.32)	(0.36)
Ambiguity Aversion	−0.07	−0.22	−1.59**	−1.87**	−0.82**	−0.60^{+}
	(0.42)	(0.52)	(0.55)	(0.64)	(0.31)	(0.34)
SVO	−0.09	0.06	−0.04	−0.15	0.04	0.04
	(0.15)	(0.09)	(0.11)	(0.33)	(0.10)	(0.09)
Spite	8.46	15.03	−8.61*	24.61**	3.53	6.17
	(7.19)	(11.38)	(3.91)	(8.85)	(3.58)	(3.78)
Female	−5.39*	−4.21	−2.01	−1.53	−0.81	−1.62
	(2.59)	(2.61)	(2.04)	(3.64)	(1.53)	(1.33)
Age	0.81	−0.35	0.45	0.11	0.21	0.34
	(0.66)	(0.31)	(0.43)	(0.26)	(0.19)	(0.23)
Constant	1.04	22.97*	13.93	−5.80	−6.13	1.63
	(16.89)	(10.79)	(20.48)	(20.53)	(11.04)	(10.29)
Treatments	3L	5L	3H	5H	All	No 5H
Other Controls	✓	✓	✓	✓	✓	✓
Realized Group Size	✓	✓	✓	✓	✓	✓
Observations	1,632	1,700	1,632	1,666	6,630	4,964
# Clusters	16	10	16	10	52	42
R^2	0.09	0.16	0.13	0.22	0.07	0.09

Note: SE clustered at group level $+\ p < 0.1$; $^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$

Heterogeneities

Figure A.16 shows effort (line 1), sabotage (line 2), received sabotage (line 3), and individual performance (line 4) depending on the rank within a group conditional on the treatments. The rank is based on the exerted effort by group for each round separately. The figure shows that there are heterogeneities between the group members. The group member that exerts the highest effort also exerts the highest sabotage. Therefore, the group member with the highest effort also receives the least sabotage of the others. This results in a high individual performance for this group member. Therefore, different to theory, less effort is destroyed, as the highest effort group member receives the least sabotage.

Figure A.16: Results per rank



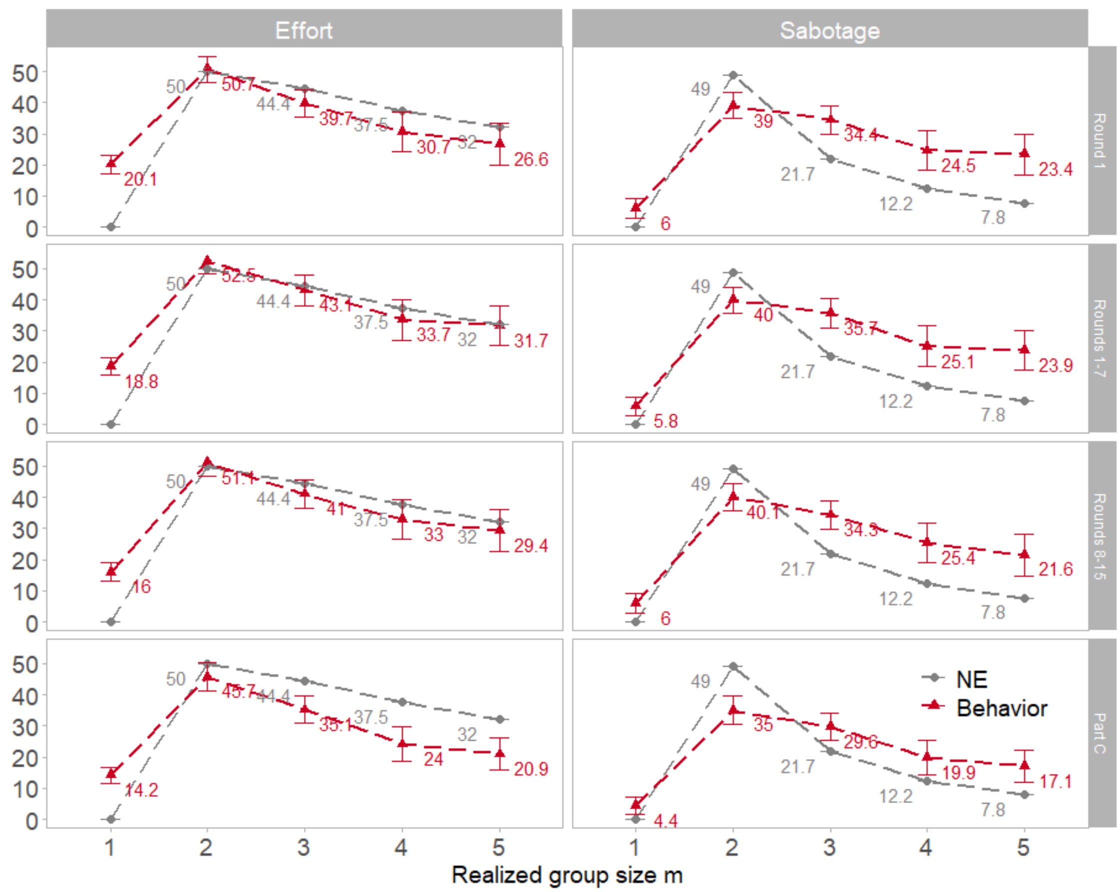
Note: The bar charts show the individual averages based on the rank by effort within a group. The x-axis shows the four treatments, whereas the y-axis in the four panels shows either effort, sabotage, received sabotage, or individual performance. The error bars show 95% confidence intervals.

A.3.2 Known Group Sizes (Group Size Disclosure)

Subsets of Rounds

Figure A.17 shows the effort and sabotage levels under disclosure for the realized group sizes conditional on a specific subset of rounds. The first line shows decisions only from the first round, the second the average from the first part of Part A, the third line from the second part of Part A, and the fourth from Part C.

Figure A.17: Robustness results average effort and sabotage levels per realized group size



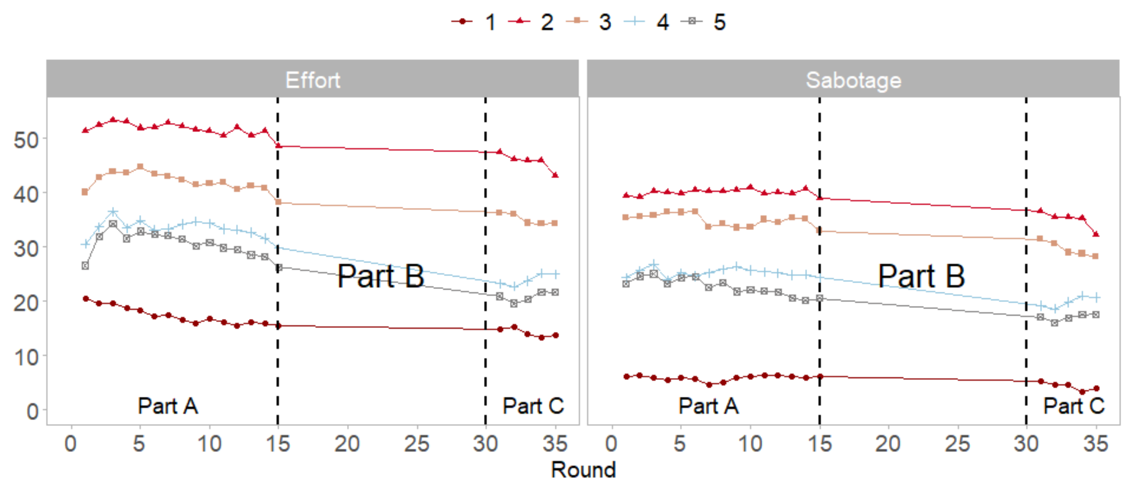
Note: The panels depict effort and sabotage levels under group size disclosure for the realized group sizes m conditional on a specific subset. Red lines show the elicited behavior, averaged over the specified subset of rounds. Blue lines show the Nash equilibrium predictions. The error bars show 95% confidence intervals.

The effort and sabotage levels are very similar between these subsets of rounds. Importantly, the significant and substantive decrease in effort and sabotage levels for an increase in the group size (except $m = 1$) is very prevalent in all of the panels. Additionally, non-parametric tests show a significant difference between the sabotage (effort) decisions for $m = 2$ and $m = 5$ at a significance level of $p < 0.001$ ($p < 0.001$) in all panels. Moreover, most of the piece-wise comparisons are statistically significant at at least $p < 0.05$.

Time Trends Part A and Part C

In this section, I analyze changes in effort and sabotage levels over time. Figure A.18 depicts these time trends for parts A and C conditional on the specific realized group size m. Overall, there is a slight decrease in effort and sabotage levels over the rounds. Importantly, the differences between the realized group sizes are not affected by the slight decrease over time – they remain relatively stable over all rounds.

Figure A.18: Time trends Part A and Part C



Note: The panels show the time trends for effort and sabotage levels in Part A and Part C. The vertical line at round 15 indicates the end of Part A and the beginning of Part C. The colors indicate the average choices for a specific group size.

Regression Results

Table A.9 shows the results of linear regressions with clustered standard errors at the matching group level, where I regress effort and sabotage on the realized group size under group size disclosure (parts A and C). I only include realized group size of $m > 1$, as being alone in the contest ($m = 1$) is a special case. I include the pre-registered controls[4] and a dummy for Part C.

Models (1) and (2) confirm the results of the main section and show a significant negative effect of the realized group size on effort and sabotage levels. It further confirms the slight time trend, as round and Part C have significant negative effects on effort and sabotage. In all models, the spite score has a significant and substantive effect on the elicited choices,

[4]The controls are: being active in the round before, having won in the round before, average sabotage and effort levels of other participants in the round before, round, determined group size in the round before, how often won in the rounds before, SVO, spite, risk, loss and ambiguity aversion, age, gender, highest degree, the field of study, the degree of concentration and understanding

showing that subjects with spiteful preferences are more competitive.

Table A.9: Linear regression effort and sabotage on realized group size based on Part A and Part C

	Dependent Variable:	
	Effort	*Sabotage*
	(1)	(2)
Realized Group Size	-6.36^{***}	-4.77^{***}
	(0.74)	(0.64)
Round	-0.47^{**}	-0.28^{*}
	(0.16)	(0.13)
Part C	-7.83^{***}	-6.58^{***}
	(1.84)	(1.67)
Risk Aversion	-1.47	-0.15
	(1.51)	(1.29)
Ambiguity Aversion	-2.48^{+}	-1.60
	(1.45)	(1.27)
Loss Aversion	-0.91	-1.59^{+}
	(1.12)	(0.87)
SVO	0.06	0.04
	(0.18)	(0.18)
Spite	18.67^{*}	29.76^{***}
	(9.07)	(8.69)
Female	9.65^{*}	5.43
	(4.47)	(4.22)
Age	-0.92	-1.78^{***}
	(0.71)	(0.51)
Constant	99.99^{***}	100.46^{***}
	(23.93)	(20.68)
Treatment Dummies	✓	✓
Other Controls	✓	✓
Observations	11,172	11,172
# Clusters	52	52
R^2	0.18	0.17

Note: Se clustered at group level $+\ p < 0.1$; $*\ p < 0.05$; $**\ p < 0.01$; $***\ p < 0.001$

A.3.3 Group Size Uncertainty

Subsets of Rounds

Tables A.10 and A.11 show average effort and sabotage levels for different subsets of rounds. The tables show that both effort and sabotage decrease when the group size increases from 3H and 5H for high enter probabilities, for all shown subsets of rounds. Additionally, non-parametric tests confirm these decreases as significant (at least $p < 0.05$) for all subsets of rounds apart from effort in the single round 1 ($p = 0.1994$). Sabotage and effort decisions are not significantly different between *3L* and *5L* in all shown subsets of rounds.

Table A.10: Effort levels under group size concealment for different rounds

Treatment	Effort NE	Elicited Effort Levels Round 1	Rounds 1-7	Rounds 8-15
3L	21.53	39.19	36.85	30.17
		(5.01)	(4.36)	(3.81)
5L	32.00	31.78	31.89	27.82
		(4.68)	(5.91)	(4.53)
3H	43.75	44.21	43.58	43.22
		(4.87)	(3.23)	(3.53)
5H	37.66	33.71	32.13	26.43
		(4.97)	(2.65)	(2.21)

Note: Average elicited effort in Part B by treatment based on different subsets, as well as the Nash equilibrium (NE). Standard errors by group and in round 1 by individual.

Table A.11: Sabotage levels under group size concealment for different rounds

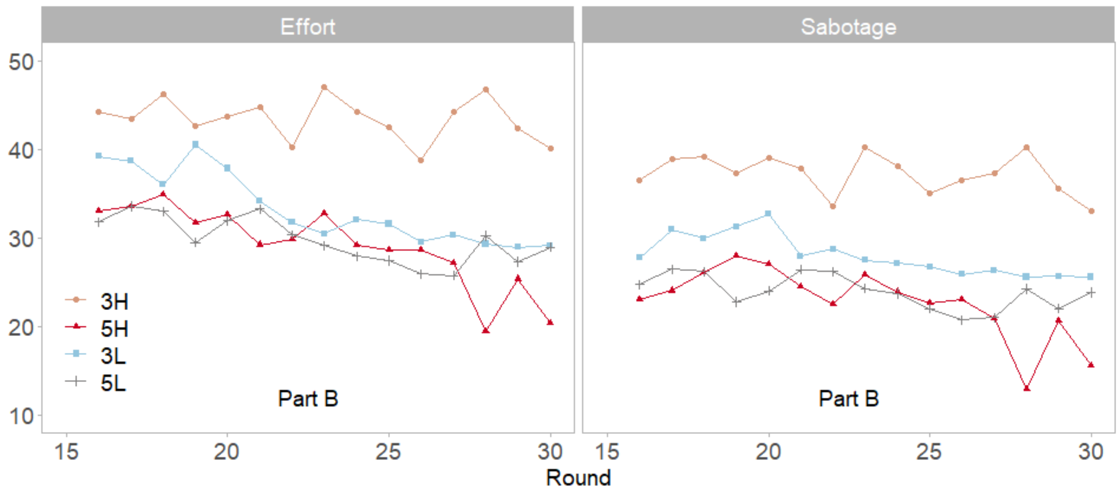
Treatment	Sabotage NE	Elicited Sabotage Levels Round 1	Rounds 1-7	Rounds 8-15
3L	19.17	27.73	29.89	26.29
		(4.09)	(4.81)	(4.03)
5L	25.50	24.74	25.25	22.73
		(4.04)	(4.09)	(3.67)
3H	30.45	36.48	37.45	36.97
		(4.47)	(3.17)	(3.62)
5H	14.36	23.55	25.04	20.67
		(4.05)	(2.87)	(2.10)

Note: Average elicited sabotage levels in Part B by treatment based on different subsets, as well as the Nash equilibrium (NE). Standard errors by group and in round 1 by individual

Time Trends Part B

Figure A.19 shows the time trends for Part B conditional on the treatments. It shows a slight decrease over time and the treatment differences remain relatively stable across the rounds.

Figure A.19: Time trends Part B



Note: The panels show the time trends for average effort and sabotage levels in Part B. The colors indicate the average choices for the specific treatment.

Regression Results

Table A.12 shows the results of a linear regression of effort and sabotage on the treatments and controls under group size uncertainty (Part B) with clustered standard errors at the matching group level. I split the sample into the treatments with a high enter probability (*3H* and *5H*) and into the treatments with a low enter probability (*3L* and *5L*) as from theory, I expect differential effects depending on the enter probabilities. Models (1) and (2) are based on a sample of treatments *3H* and *5H*, and models (3) and (4) of *3L* and *5L*. Furthermore, I include the pre-registered controls.[5]

The models confirm the results from the main section. In models (1) and (2), I find a significant decrease in effort and sabotage for Treatment *5H* compared to Treatment *3H*. In models (3) and (4), I do not find any significant effect of Treatment *5L* compared to *3L*.

Furthermore, I confirm the negative time trend, as the round variable has a significant negative effect on effort and sabotage in all models. Under high enter probabilities, I find

[5]The controls are: being active in the round before, having won in the round before, average sabotage and effort levels of other participants in the round before, round, determined group size in the round before, how often won in the rounds before, SVO, spite, risk, loss and ambiguity aversion, age, gender, highest degree, the field of study, the degree of concentration and understanding

a significant negative correlation between loss aversion and effort and sabotage, and under low enter probabilities, I find a significant negative correlation between ambiguity aversion and effort and sabotage levels.

Table A.12: Linear regression effort and sabotage on treatments under group size uncertainty

	Dependent Variable:			
	Effort	*Sabotage*	*Effort*	*Sabotage*
	(1)	(2)	(3)	(4)
Round	-0.94^{***}	-0.72^{**}	-0.83^{**}	-0.40^{*}
	(0.28)	(0.24)	(0.29)	(0.19)
Treatment 5H	-13.49^{***}	-10.76^{*}		
	(3.94)	(5.33)		
Treatment 5L			-3.18	3.55
			(5.90)	(3.35)
Risk Aversion	-1.12	-0.27	-1.15	-3.14^{*}
	(2.02)	(2.19)	(1.71)	(1.36)
Ambiguity Aversion	-0.72	-0.20	-3.88^{*}	-3.03^{**}
	(1.89)	(1.91)	(1.86)	(1.08)
Loss Aversion	-3.82^{**}	-4.11^{***}	1.64	1.16
	(1.22)	(1.18)	(1.55)	(1.49)
SVO	0.36	0.29	0.09	0.13
	(0.31)	(0.32)	(0.27)	(0.24)
Spite	24.33	30.94^{+}	10.99	3.53
	(18.54)	(16.31)	(13.88)	(11.62)
Female	12.12^{**}	8.63^{+}	2.72	5.21
	(4.63)	(5.07)	(6.76)	(6.30)
Age	-0.41	-0.71	0.18	-0.29
	(0.94)	(0.80)	(1.30)	(1.07)
Constant	66.61^{*}	74.67^{**}	3.55	35.55
	(31.44)	(25.13)	(37.69)	(33.99)
Treatments	3H, 5H	3H, 5H	3L, 5L	3L, 5L
Other Controls	✓	✓	✓	✓
Observations	1,455	1,455	1,470	1,470
# Clusters	52	52	52	52
R^2	0.24	0.21	0.17	0.23

Note: SE clustered at group level $+\ p < 0.1$; $*\ p < 0.05$; $**\ p < 0.01$; $***\ p < 0.001$

A.4 Instructions

In the following, I show the instructions used in this experiment.

A.4.1 Tutorial

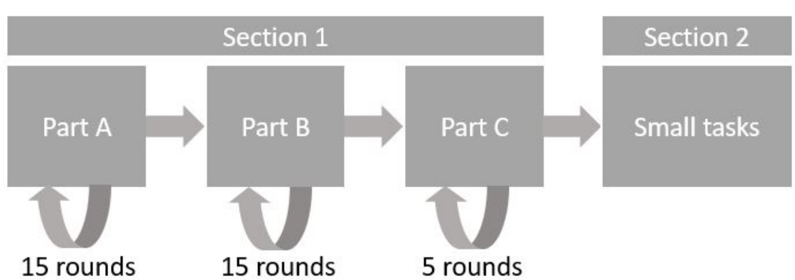
Welcome!

Welcome to this experiment, and thank you for your participation. In this experiment, you have the opportunity to earn points. The number of points depends on your decisions, the decisions of the other participants of this experiment, and luck. After the experiment is finished, we will translate the number of points into Euros at an exchange rate of 100 points = 9 Euros (1 point = 9 Euro cents). On top of that, you will receive 1 Euro.

One note before we begin: It is very important to us, that all participants stay concentrated only on the experiment. If you have a question during the experiment, please write to us in the Zoom chat.

Procedure

The experiment consists of two sections. Section 1 is the main part of this experiment, where you will interact with other participants of this experiment. In section 2, you will complete several small tasks. Section 1 consists of part A, part B, and part C. In total, there will be 35 rounds, as shown in the picture.



At the end of the experiment, 3 of the 35 rounds will be randomly chosen for your payment. The average earnings of these 3 selected rounds determine your payoff of section 1. **Hence, any of the rounds can be payoff-relevant for you. Therefore, it is advisable to think about each decision carefully.**

Additionally, we will donate to one of the following five charities. We will explain in the following pages, how the amount will be determined. One charity will be randomly determined by the computer at the end of the experiment. The average donation of the 3 out of 35 randomly selected rounds is taken.

- Amnesty International
- Doctors Without Borders
- German Red Cross
- Greenpeace
- UNICEF

We start with a tutorial for part A. Please go through the tutorial attentively and contact us in the Zoom chat in case you have any question. At the end of the tutorial, the computer will ask you several comprehension questions about the instructions. After that, part A will begin.

Part A - Tutorial 1/5

Welcome to the tutorial of part A. The tutorial will prepare you step-by-step for the actual experiment. We start with the most simple version and successively add layers.

Opportunity to win 200 points

You are grouped with one other participant of this experiment. One of you will win a prize of 200 points.

Option A

You will begin each round with a start balance of 200 points. You can choose to keep these points or invest some of them in Option A. The maximum you can invest is 100 points. Any number invested in Option A increases your 'performance'. The other group member can also invest points in Option A to increase his/her 'performance'. The higher your performance is in comparison to the other group member's performance, the higher is your probability to win the 200 point prize. Your and the other group member's performance and probability to win are calculated as follows:

$$\text{Performance} = \text{Points invested in Option A ('Option-A points')}$$

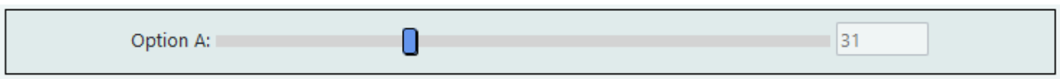
$$\text{Your probability to win} = \frac{\text{Your performance}}{\text{Your performance} + \text{Other group member's performance}}$$

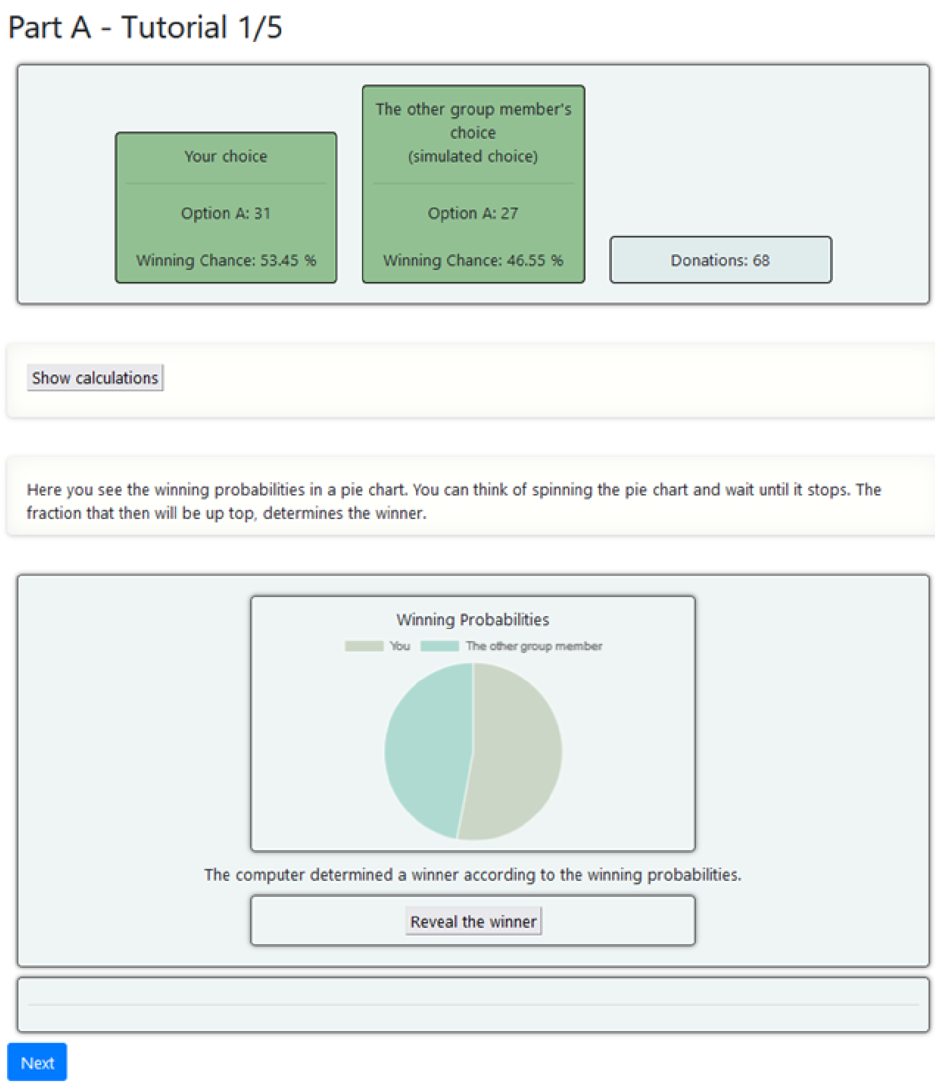
If both performances are 0, the winning probability is 50% for each group member. Your and the other group member's performance not only influence the winning probabilities. We will also donate money to a charity depending on the total performance. The donations are calculated as follows:

$$\text{Donations} = \text{Your performance} + \text{Other group member's performance} + 10$$

TRY IT OUT!

Please choose how much to invest in Option A. Any points that you don't invest (out of the 200) are yours to keep. The computer will simulate a random choice for the other group member.





Part A - Tutorial 2/5

Option B

In the actual experiment, you will have a second option. Additionally to Option A, you can invest up to 100 points into Option B. With Option B you decrease the other group member's performance. Likewise, the other group member can decrease your performance by investing into Option B. Your and the other group member's performance and probability to win are then calculated as follows. It is not important to remember this formula or to fully understand it. We show it for full transparency. For the actual experiment, you will have access to a calculator which helps you get a sense of how the choices affect your performance and probability to win.

$$\text{Your performance} = \frac{\text{Your Option-A points}}{1 + \text{The other group member's Option-B points}}$$

$$\text{Your probability to win} = \frac{\text{Your performance}}{\text{Your performance} + \text{The other group member's performance}}$$

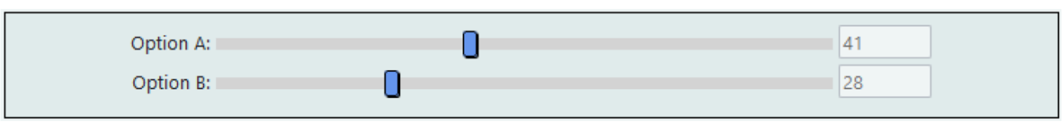
If both performances are 0, the winning probability is 50% for each group member. The donations are calculated as before:

$$\text{Donations} = \text{Your performance} + \text{Other group member's performance} + 10$$

Hence, Option A increases the donations and Option B decreases the donations.

TRY IT OUT!

Please choose how much to invest in Option A and Option B. Any points that you don't invest (out of the 200) are yours to keep. The computer will simulate random choices for the other group member.



[[*Feedback shown similar to before*]]

Part A Tutorial 3/5

Your Group

In the actual experiment, instead of one other participant, you will be paired with 4 other participants of this experiment. You stay in this group until the end of the first section of this experiment.

Now, one of the group members will win 200 points. Every group member can invest into Option A and Option B. As before, Option A increases the own performance. Option B decreases the performance of **all other group members simultaneously**. Your and the other group members' performances and probabilities to win are calculated as follows:

$$\text{Your performance} = \frac{\text{Your Option-A points}}{1 + \text{All other group members' Option-B points}}$$

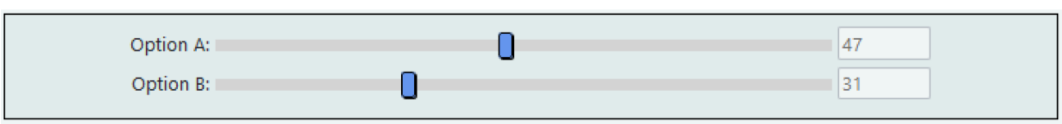
$$\text{Your probability to win} = \frac{\text{Your performance}}{\text{Your performance} + \text{All other group members' performances}}$$

If all performances are 0, the winning probabilities are the same for all group members (20% for every group member). The donations are calculated as before:

$$\text{Donations per group} = \text{Sum of all performances} + 10$$

TRY IT OUT!

Please choose how much to invest in Option A and Option B and then press on next. The computer will simulate random choices for all other group members.



[[*Feedback shown similar to before*]]

Part A - Tutorial 4/5 *Active and inactive group members*

In the actual experiment, not every group member will be active in each round. In every round, the computer will randomly choose a number of active group members between 1 and 5 and determines randomly who they are.

If you are inactive...

... you do not interact with anyone in this round.

If you are active...

... you will interact with all other active group members. One of the active group members will win the 200-point prize. Depending on the random choice of the computer, you interact with either 0, 1, 2, 3, or 4 other active group members.

Procedure

In every round, you make your decisions before the computer determines whether you are active or not. We will ask you how many points you would want to invest if

- you are the only active group member
- there is 1 other active group member
- there are 2 other active group members
- there are 3 other active group members
- there are 4 other active group members

After that, the computer will randomly determine the number of active group members. The chances of the different numbers of active group members are not even. As you make your decisions depending on the specific number of other active group members, the chances are not further important for part A. The chance of being active is NOT influenced by the choices.

If you are active, the computer will implement your decisions for the specific number of active players.

If you are the only active group member, you will win the prize of 200 points for sure, independent of how much you invest into Option A and Option B. However, you still have to pay for your investment choices.

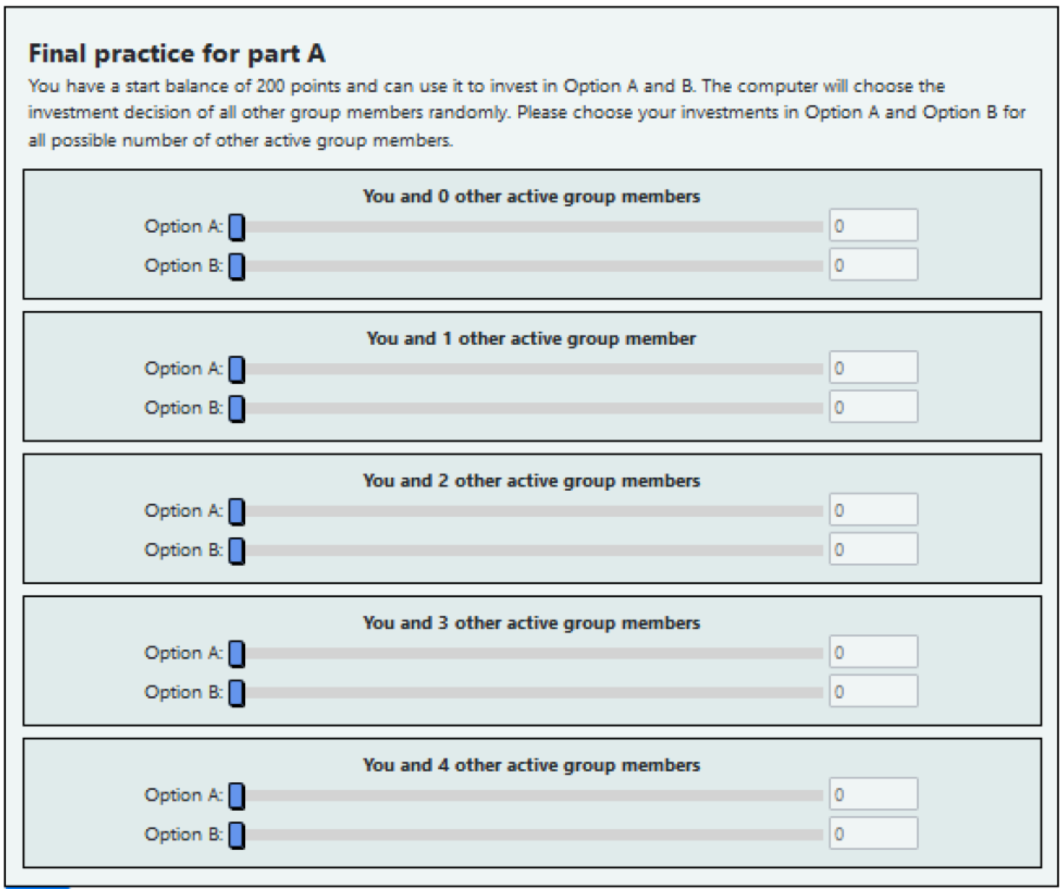
If you are inactive, your choices do not matter neither for the performances and winning probabilities, nor for the donations. You will not have to pay for your decisions.

Payoff

Your payoff in this round is determined as follows:

If you are inactive:	**If you are active and win:**	**If you are active and lose:**
+200 (start balance)	+200 (start balance)	+200 (start balance)
	+200 (winning prize)	+0 (no prize)
	- points you invested	- points you invested
200	400 - points you invested	200 - points you invested

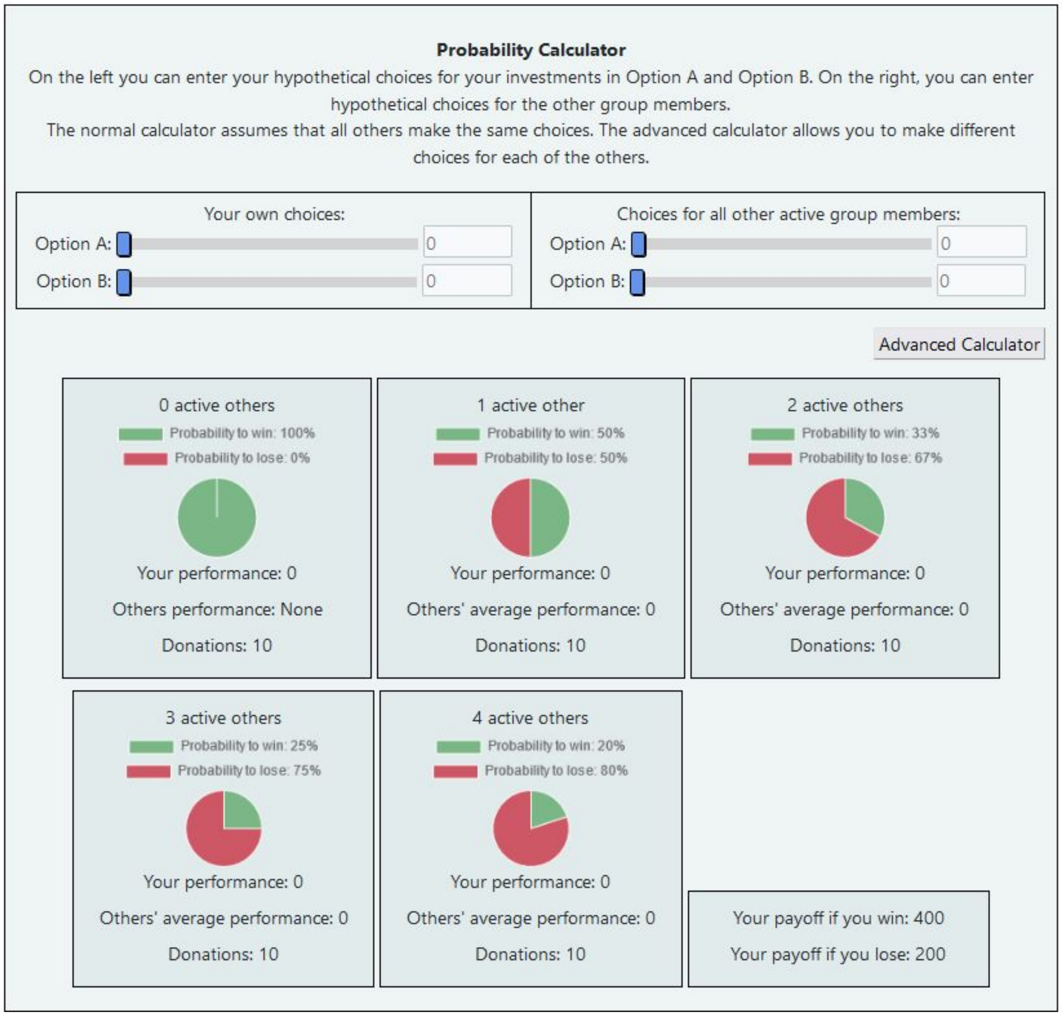
There will be a time limit for your decisions. In the first rounds you will have more time than in later rounds.



[[*Feedback shown similar to before*]]

Part A - Tutorial 5/5

You will have access to a calculator throughout the entire experiment. The probability calculator shows you how your choices and the other active group members' choices affect your probability to win and the donations. Please take your time to familiarize yourself with the underlying mechanisms.



Just for your information, you will always be able to access a summary of the instructions by scrolling down.

Summary of Instructions

1. Everyone decides how many points of the start balance to invest in Option A and Option B.

$$\textbf{Your performance} = \frac{\textbf{Your Option-A points}}{\textbf{1 + All other group members' Option-B points}}$$

- Option A increases your own performance
- Option B decreases all other active group members' performances simultaneously

$$\textbf{Your probability to win} = \frac{\textbf{Your performance}}{\textbf{Your performance + All other group members' performances}}$$

- The higher your performance in comparison to the other active group members' performances, the higher your probability to win

The performances of all active group members influence the donations:

$$\textbf{Donations per group} = \textbf{Sum of all performances} + \textbf{10}$$

- The higher the performances, the higher the donations
- Option A increases the donations
- Option B decreases the donations

2. The computer determines a number of active group members between 1 and 5 and determines the active group members. Everyone has the same probability to become active.

3. The computer determines the winning probabilities with the choices of the active group members

- The choices of all active group members for the specific number of active group members are picked.
- The performances and winning probabilities are calculated based on the choices of the active group member.
- The donations are calculated with the performance of every active group member.

4. The computer will determine the winner according to the winning probabilities. The winner receives the 200-point prize. If you are inactive, your choices do not count.
Your individual payoff is determined as follows:

If you are inactive:	**If you are active and win:**	**If you are active and lose:**
+200 (start balance)	+200 (start balance)	+200 (start balance)
	+200 (winning prize)	+0 (no prize)
	- points you invested	- points you invested
200	400 - points you invested	200 - points you invested

Quiz

Here is a little quiz. After you have answered all quiz questions correctly (you have several tries), we can begin with part A. Remember, you can always scroll down to see an overview of the instructions.

Q1: How many participants will be in your group, including you?

- 1
- 2
- 3
- 4
- 5

Q2: Provided that you are active, how can you increase your probability to win?

- Increase my performance and reduce the other active group members' performances
- I can't
- Increase the other active group members' performances

Q3: What can you do with Option A?

- Increase my performance
- Increase the other active group members' performances
- Decrease the other active group members' performances

Q4: What can you do with Option B?

- Decrease my performance
- Increase the other active group members' performances
- Decrease all other active group members' performances simultaneously
- Decrease the performance of another active group member of my choice

Q5: How can you increase the donations?

- Increase my performance (by investing in Option A)
- Decrease my performance (by investing less in Option A)

Q6: How can you decrease the donations?

- Increase my performance (by investing in Option A)
- Decrease the other active group members' performances (by investing in Option B)

Q7: Suppose that you are active and that a round was selected for payment. Who is affected by your decisions?

- Me and the charity
- Me, the other active group members of my group, and the charity
- Everyone of my group and the charity

Q8: How many group members are active in one round?

- 3
- 5
- This is determined randomly in every round
- This is determined randomly in the first round and stays the same until the end of the experiment

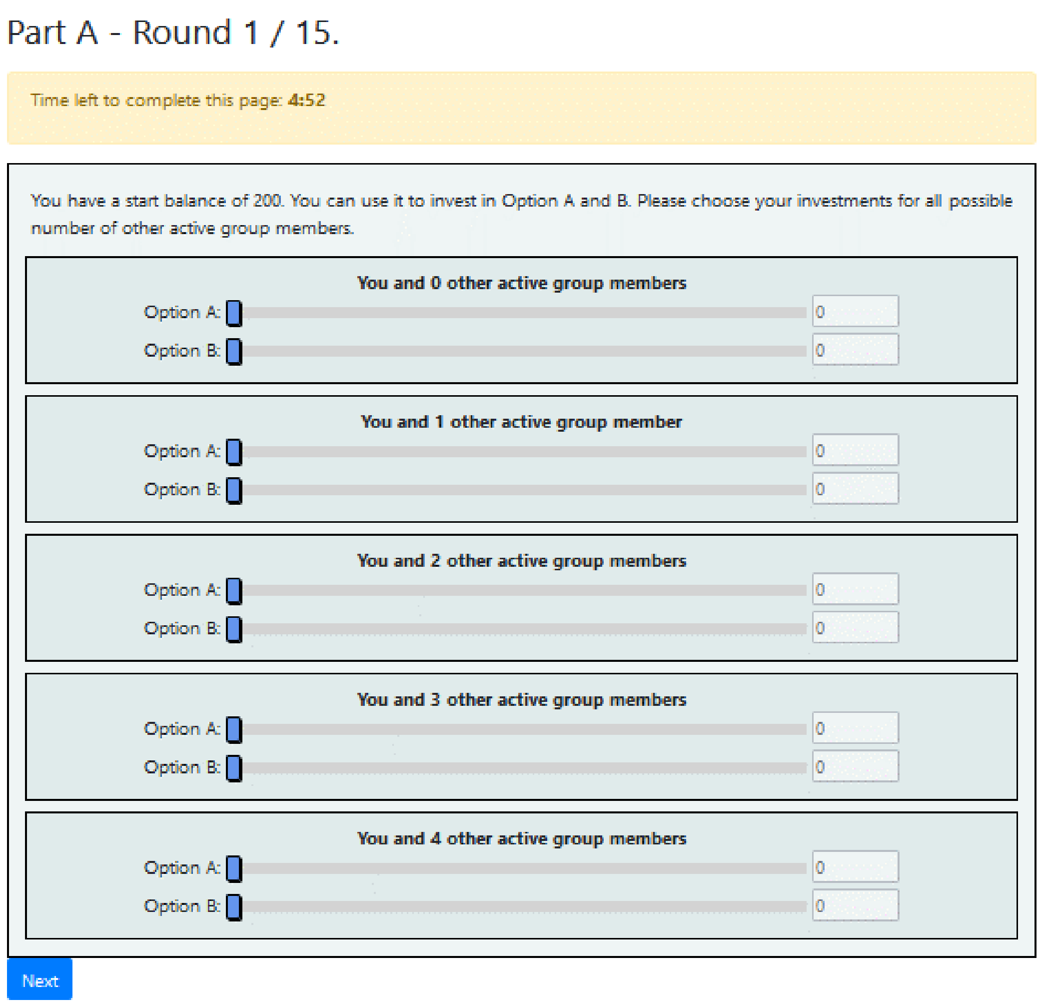
Q9: What happens if only one group member becomes active?

- This group member wins the 200-point prize independently of his/her choices
- There will never be just one active person
- There is a 50

Q10: In each round, you receive a start balance of 200 points. What happens with the points that you do not invest in Option A or Option B?

- These points are destroyed and will not be added to my payoff of this round
- I can keep the points and they will be added to my payoff of this round

A.4.2 Section 1 - Part A



Part A - Round 1 / 15.

Time left to complete this page: **4:52**

You have a start balance of 200. You can use it to invest in Option A and B. Please choose your investments for all possible number of other active group members.

You and 0 other active group members

Option A: 0

Option B: 0

You and 1 other active group member

Option A: 0

Option B: 0

You and 2 other active group members

Option A: 0

Option B: 0

You and 3 other active group members

Option A: 0

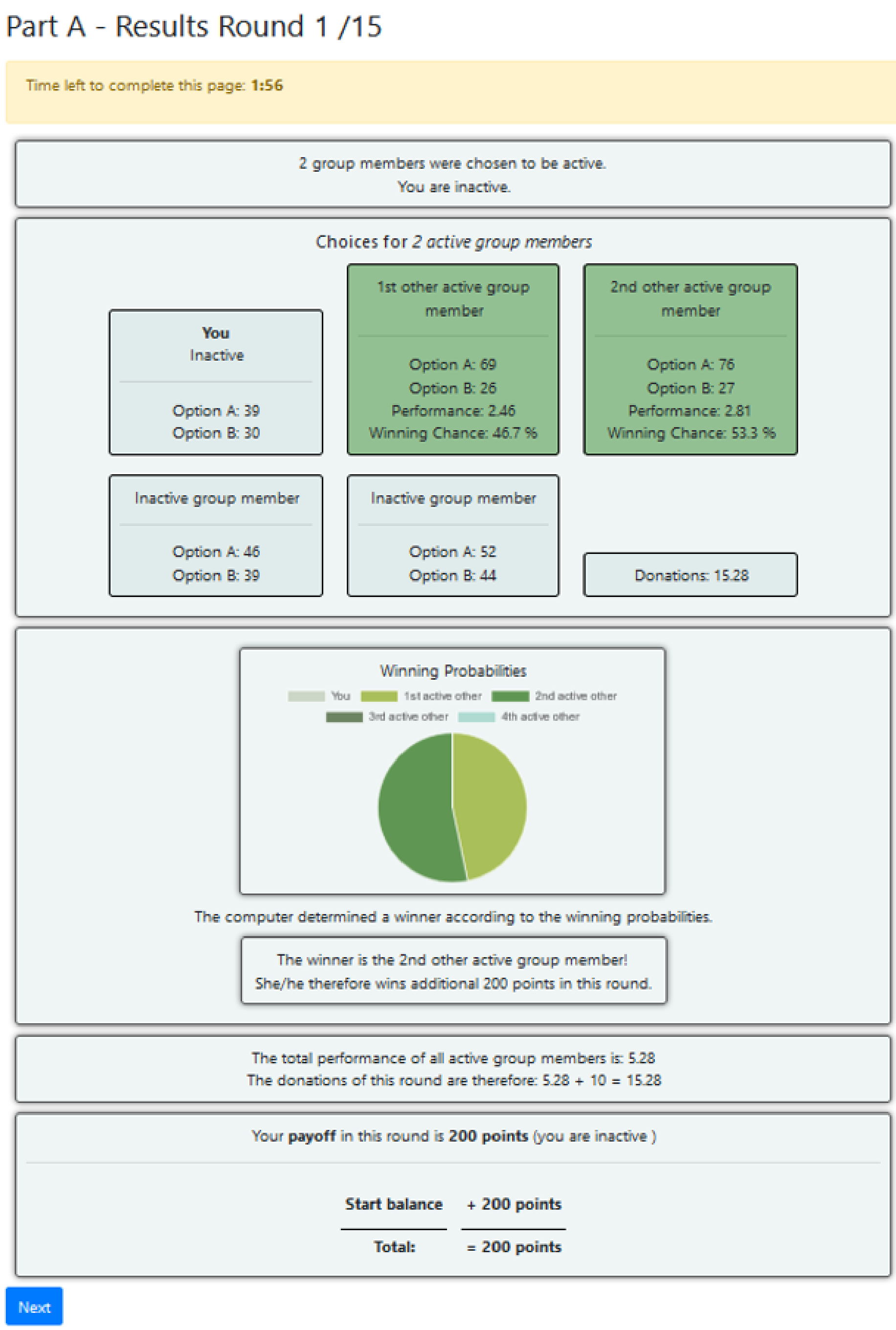
Option B: 0

You and 4 other active group members

Option A: 0

Option B: 0

Next

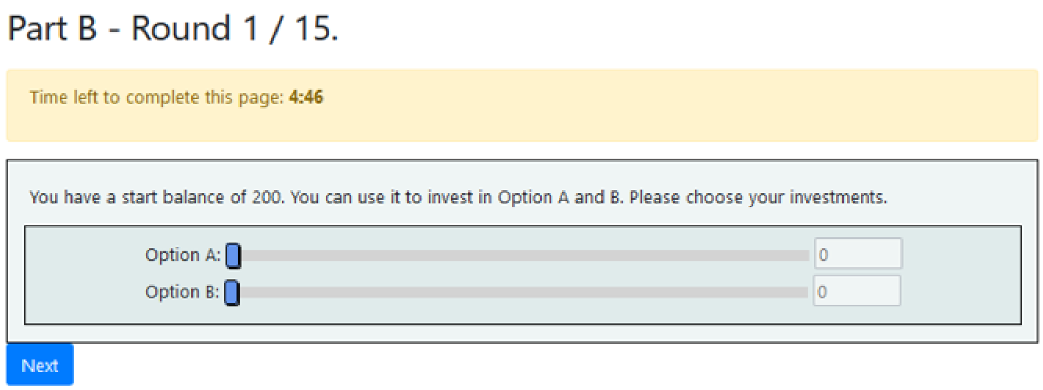


A.4.3 Section 1 - Part B

Part B is very similar to Part A. You stay in the same group as in part A. The only difference to part A is that we do not ask you for your decisions for every possible number of other active group members. Instead, we ask you for **one** decision for Option A and for **one** decision for Option B. In other words, you only decide once for Option A and once for Option B, and this one decision each has to fit all possible scenarios (0 others, 1 other, 2 others, 3 others, 4 others). Therefore, it is advisable that you think about how likely these scenarios are and adjust your decisions accordingly.

The following table and pie chart show the probabilities for the number of other active group members in each round, given that you are active.

Number of other active group members	0	1	2	3	4
Probability of Occurrence	<1%	5%	21%	42%	32%



Part B - Results Round 1 /15

Time left to complete this page: **1:50**

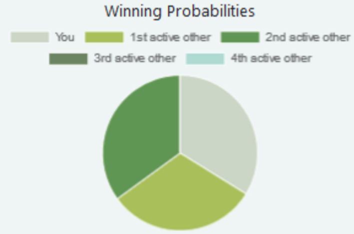
3 group members were chosen to be active.
You are active.

Choices

You Active	1st other active group member	2nd other active group member
Option A: 48 Option B: 32 Performance: 1.07 Winning Chance: 33.8 %	Option A: 62 Option B: 14 Performance: 0.98 Winning Chance: 31.2 %	Option A: 52 Option B: 30 Performance: 1.11 Winning Chance: 35.0 %

Inactive group member	Inactive group member	
Option A: 56 Option B: 38	Option A: 54 Option B: 37	Donations: 13.16

Winning Probabilities



The computer determined a winner according to the winning probabilities.

The winner is the 1st other active group member!
She/He therefore wins additional 200 points in this round.

The total performance of all active group members is: 3.16
The donations of this round are therefore: 3.16 + 10 = 13.16

Your **payoff** in this round is **120 points** (you are active)

Start balance	+ **200 points**
Costs Option A	- **48 points**
Costs Option B	- **32 points**
Total:	= **120 points**

Next

A.4.4 Section 1 - Part C

The rules for part C are identical to the rules of part A. Part Consists of 5 rounds. You remain in the same group as before.

[[Elicitation and Feedback the same as in Part A.]]

A.4.5 Section 2

In this final section 2, we continue with several small tasks. Unlike in section 1, we do not use points anymore. Instead, you will be deciding about Euro cents.

In the following task, you will be randomly paired with another participant. You will be making a series of decisions about allocating cents between you and this other person. You will make 9 choices. The other person also makes the same 9 choices. It will be randomly determined whether your choices or the other person's choices constitute the payoff for the two of you. 1 of the 9 choices will be randomly picked at the end of the experiment.

Table A.13: Section 2 - Choices

You receive	85	85	85	85	85	85	85	85	85
	○	○	○	○	○	○	○	○	○
Other receives	85	76	68	59	50	41	32	24	15
You receive	85	87	89	91	92	94	96	98	100
	○	○	○	○	○	○	○	○	○
Other receives	15	19	24	28	32	37	41	46	50
You receive	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	○
Other receives	100	98	96	94	92	91	89	87	85
You receive	50	54	59	63	68	72	76	81	85
	○	○	○	○	○	○	○	○	○
Other receives	100	89	79	68	58	47	36	26	15
You receive	100	94	88	81	75	69	62	56	50
	○	○	○	○	○	○	○	○	○
Other receives	50	56	62	69	75	81	88	94	100
You receive	100	98	96	94	92	91	89	87	85
	○	○	○	○	○	○	○	○	○
Other receives	50	54	59	63	68	72	76	81	85
You receive	70	70	70	70	70	70	70	70	70
	○	○	○	○	○	○	○	○	○
Other receives	100	98	96	94	92	91	89	87	85
You receive	70	68	65	62	60	58	55	52	50
	○	○	○	○	○	○	○	○	○
Other receives	100	96	92	89	85	81	78	74	70
You receive	100	100	100	100	100	100	100	100	100
	○	○	○	○	○	○	○	○	○
Other receives	100	98	96	94	92	91	89	87	85

The following three pages each present 10 scenarios for which you should make a decision. In each case, you decide between a lottery and a fixed payment.

After you have made all choices, one of the pages and one of the scenarios will be randomly chosen for your payment. If you have chosen the fixed payment, you will get the corresponding payoff for sure. If you have chosen the lottery, it will be randomly determined (according to the corresponding probabilities) whether you receive the low or high outcome.

Table A.14: Section 2 - Page [[1 or 2]]

Please choose for every row, whether you prefer the lottery or the fixed payment

Lottery			Fixed payment
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 5 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 10 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 15 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 20 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 25 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 30 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 35 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 40 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 45 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	winning 50 cents for sure

Table A.15: Section 2 - Page [[1 or 2]]

Please choose for every row, whether you prefer the lottery or the fixed payment

Lottery			Fixed payment
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 5 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 10 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 15 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 20 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 25 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 30 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 35 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 40 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 45 cents for sure
50% prob. of winning 0 cents, 50% prob. of winning 50 cents	○	○	losing 50 cents for sure

Table A.16: Section 2 - Page 3

For every row, please choose whether you prefer the lottery or the fixed payment. In the lottery, **p denotes the probability in percent with which you lose.** The computer will randomly determine this probability after your decisions. p can be between 0 and 100.

Lottery			Fixed payment
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 5 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 10 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 15 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 20 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 25 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 30 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 35 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 40 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 45 cents for sure
p% prob. of winning 0 cents, (100-p)% prob. of winning 50 cents	○	○	winning 50 cents for sure

Final questionnaire

Lastly, please enter the following information.

Your age: ______________

Please indicate the gender you most identify with:

- female
- male
- other

Please indicate your field of study: ______________

In which semester are you? ______________

Please indicate your highest degree:

- HighSchoolDegree
- Bachelor
- Master
- PhD
- Other

How concentrated were you during the experiment?

- 1 - not at all
- 2 - little concentrated
- 3 - medium
- 4 - mostly concentrated
- 5 - very concentrated

How well did you understand the experiment?

- 1 - not at all
- 2 - not well
- 3 - medium
- 4 - mostly understood it
- 5 - understood it well

Your Payoff

Thank you for participating in this experiment. You earned a total of **XY Euros. IMPORTANT!** Please write the following payment code on the formula that you received by E-Mail. Without this code, the payment can not be made: XYZ123456

[[*Calculations of payoff were shown.*]]

Chapter 2

Spite in Litigation

with Wladislaw Mill

2.1 Introduction

It is well known that some plaintiffs sue defendants not only to seek justice but also out of malice, spite, and pure anger. Malicious and spiteful litigants are suing and going to court just to harm and punish the opponent. They derive utility from the harm inflicted upon others, either because they are inherently spiteful or because the situation triggers spitefulness.

Such malicious litigation is a very popular and regularly recurring theme in TV shows about law and medicine. Yet, this pattern is not only fictional but also has a very real match in legal practice: the 'Vexatious Litigant', which is typically defined as follows:

> '[...]Vexatious litigation is meant to bother, embarrass, or cause legal expenses to the defendant.[...]'[1]

Vexatious litigants are people who go to court to harm and bother the defendant. Very often, they file frivolous lawsuits. Frivolous lawsuits – lawsuits typically filed by a party who is aware that the case is without merit – waste time, money, and judicial resources.[2] Subjects who repeatedly engage in vexatious litigation might, in some jurisdictions, be added to lists of vexatious litigants. In Great Britain, for example, this means that one is forbidden from starting a civil case without court permission.[3]

[1] See Legal Information Institute (2018).

[2] See the argument in Anderson (1997), Post (2011), and Yago (1999).

[3] For a list of vexatious litigants in Great Britain see `https://www.gov.uk/guidance/vexatious-litigants`.

Malicious and spiteful litigation is not uncommon[4] and often found for example between disputing neighbors, alienated partners, angry siblings, and business rivals.[5] Divorces and malpractice suits are particularly prone to malicious litigation.[6] But malicious litigation can also occur between mere acquaintances.[7] These examples underline that spiteful litigation can occur either because agents are inherently spiteful, or because their spitefulness is triggered by the situation, such as one party perceiving the other as unreasonable, or by the behavior of the other party that led to litigation in the first place. Excessive litigation expenditures that come with such spitefulness do not only waste resources for the litigant, they also force the defendant to increase legal expenses to maintain the same chances of winning the case. Therefore, the question arises: How can a legal system be designed to decrease such wasteful litigation behavior driven by spitefulness?

One such possible way is the choice of the fee-shifting rule, which determines who (the defendant or the plaintiff) has to pay for whose legal costs. Which of the rules should be implemented remains an important question that is still discussed today. Lawyers, as well as judges, create substantial costs in the litigation process and it seems plausible that the loser of litigation should at least pay for her own lawyers. Therefore, the core question of the fee-shifting debate is whether and how much the loser has to pay for the winner's legal costs.

Two common approaches are typically discussed in the literature: the American and the English fee-shifting rule. Under the American rule, everybody has to pay their own expenses. Hence, under this rule, there are no additional costs of losing. Under the English rule, the loser has to pay the legal expenditures of the winner – up to a certain amount. This way, frivolous lawsuits are hoped to be discouraged. In the theoretical literature, it is argued that

[4]Similar arguments are made by Chen and Rodrigues-Neto (2023); Guha (2016); Kisner (1976); Philippi (1983).

[5]There are several examples of malicious litigation among many others: Singleton v Singleton, 68 Cal. App. 2nd , 699 (1945) represents a case of malicious litigation between siblings; GRAHAM v. GRIFFIN, 66 Cal. App. 2nd, 116 (1944) is a case of malicious litigation between neighbors. Singleton v Perry, 45 Cal. 2nd 492 (1955) and more so Davey v. Dolan, 453 F. Supp. 2d 749 (2006) show cases of spiteful litigation towards estranged partners and their families. Crowley v. Katleman, Cal. P. 2nd 1083 (1994) presents a case of former friends engaging in malicious litigation. CSC (Contemporary Services Corporation) v Staff Pro Inc, 152 Cal. App. 4th 1043 (2007) shows a fascinating case of several rounds of malicious litigation. Silver v. Gold, 211 Cal. App. 3d, 17 (1989) shows a case of malicious prosecution between business rivals and Cantu v. Resolution Trust Corp., 4 Cal. App. 4th, 857 (1992) and Casa Herrera, Inc. v. Beydoun, 32 Cal. 4th, 336 (2004) show cases of malicious prosecution of people in business. Bertero v. National General Corp. , 13 Cal. 3d, 43 (1974) depicts a case of malicious prosecution of employer and employee.

[6]See Kisner (1976) and Philippi (1983). One such example of a malicious malpractice suit is Lackner v Lacroix, 25 Cal 3rd, 747 (1979). Many more examples can be found in Kisner (1976) footnote 11 and footnote 8. See also Philippi (1983) footnote 6 for several examples of malicious litigation in malpractice suits and footnote 11 refers to a study arguing that most medical malpractice suits are without merit.

[7]For example, Drainville v. Vilchez, 2014 ONSC 4060 (CanLII) presents a case of malicious litigation between two truck drivers in Canada.

the English rule reduces the number of lawsuits filed by plaintiffs with such low-merit cases, with the downside that the total number of lawsuits increases compared to the American rule (see Spier, 2007).

It is not obvious how spiteful preferences impact litigation behavior under each rule. At first glance, one might think that spite simply increases litigation expenses. However, it is plausible that this depends on the merit of the case and on the fee-shifting rule. Under the American rule, there is a trade-off between harming the opponent and harming oneself as for very high-merit cases any additional dollar has to be bared by the 'attacker' and does not improve own winning chances substantially (similarly under very low-merit cases). Under the English rule, one might want to harm the other especially if the merit is high (as the chances of harming are high), but for low merit, any additional expense would almost inevitably backfire as winning chances are low. Hence, spite might have differential effects depending on the merit and the fee-shifting rule.

The goal of this paper is to study how litigation predictions of the two fee-shifting rules change if agents are not purely self-interested but also motivated by spiteful preferences. We go one step further and additionally ask how spite affects pre-trial settlement requests under the shadow of (spiteful) litigation. Finally, we address whether one of the rules is better suited to protect agents from the harm caused by excessive spiteful litigation expenditures. To answer these questions, we first build a theoretical model to derive predictions, which we then test with the help of an experiment.

Our theory, which specifically accounts for spiteful preferences, predicts that litigation expenditures are overall higher under the English rule compared to the American – but with lower expenditures for frivolous low-merit cases under the English rule. Litigation expenditures are overall higher for spiteful agents compared to non-spiteful agents because they receive additional utility from winning and additional disutility from losing. This spite-induced increase is the same between the two fee-shifting rules. However, there are differential effects depending on the merit of the case. Under the American rule, there is a proportional increase for all merit levels, whereas under the English rule, spite increases a specific range of low-merit cases only. This is because the English rule incentivizes agents to either spend all resources up to the value of the winner's prize or none at all. Consequently, spite can only increase those (low-merit) cases, where agents do not spend all of their resources yet – but only up to a specific merit. That is because, in cases with almost no merit, any additional litigation expense would not sufficiently increase winning chances to compensate for the additional costs. We also show that spiteful preferences only affect settlement behavior under the American but not under the English rule. This distinction arises from the interplay of two spite-induced countervailing forces. First, spite influences bargaining power by changing the expected litigation outcome and associated utility. Second, spite affects the utility from the bargaining outcome itself, as spiteful players receive additional disutility from any concession to their opponent. Under the American rule, the force from the expected litigation utility is more pronounced, but perfectly attenuated under the English rule by the force from the settlement stage.

To test these theoretical predictions, we run an experiment, where we can exogenously vary not only the fee-shifting rule but also the merit of the case. The controlled environment of an experiment also allows us to elicit both settlement requests and litigation expenditures for all participants, and hence, to shut down the selection effect for litigation expenditures. Additionally, we can explicitly measure subjects' spitefulness via two different measures.

With our results, we can confirm part of the theoretical predictions. In particular, we find that first, the English rule leads to overall higher expenditures for all merit levels, including low-merit cases. Unlike theory would predict, we do not find that the English rule discourages frivolous (low-merit) lawsuits. Instead, we find that the English rule encourages any kind of lawsuits, including frivolous ones. Concerning settlement requests, we do not find any significant differences between the two rules, leading to the same litigation probabilities across the two fee-shifting rules.

Second, subjects exhibiting more spiteful preferences spend more on litigation and request higher settlement amounts than those with lower spiteful preferences under both rules. This increase in litigation expenditures is more pronounced under the American compared to the English fee-shifting rule, driven by a constant increase for all merit levels under the American rule compared to an increase for low-merit cases only under the English rule, as our predictions suggest. There is no such differential effect depending on the rule on settlement requests. Consequently, the English rule seems to be more robust towards spiteful preferences, driven, however, by overall higher litigation expenditures for less spiteful players.

Third and finally, we show that being spiteful does not pay off, as the expected payoff is lower for more spiteful subjects independent of being matched only with either more or less spiteful subjects. This decrease is more pronounced under the American rule. The harm of being matched with a spiteful player, however, is similar across the two rules. As a consequence, the English rule can help to protect spiteful players from the monetary harm they inflict on themselves.[8] The harm they inflict upon others, however, can not be decreased by the choice of the fee-shifting rule.

The contribution of this paper is threefold. First, we contribute to the empirical (Snyder and Hughes, 1990; Hughes and Snyder, 1995; Fenn et al., 2017; Helland and Yoon, 2017; Helmers et al., 2021) and experimental literature (Main and Park, 2000, 2002; Inglis et al.,

[8]The monetary costs, however, may be offset by spite-induced utilities.

2005; Gabuthy et al., 2021; Massenot et al., 2021) that study the differences between the two fee-shifting rules for litigation or settlement behavior. While the existing literature suggests that the English rule leads to higher litigation expenditures overall, this paper is the first to provide (experimental) evidence of the influence of the merit of the case under the two fee-shifting rules for litigation and pre-trial bargaining.[9] Importantly, we shut down the selection effect of bargaining on litigation and therefore can speak directly to the influence of the fee-shifting rule and merit on litigation expenditures. We show that the English rule leads to higher litigation expenditures for *all* merit levels compared to the American rule, while there is no difference in settlement requests. Hence, the English rule seems not to be better suited to deter frivolous low-merit lawsuits, on the contrary, it even leads to higher expenditures for such lawsuits.

Second and most importantly, we provide consistent evidence that litigation and settlement behavior is sensitive to spiteful preferences and that this effect depends on the fee-shifting rule. The current experimental investigations of the fee-shifting rules do not account for spiteful preferences and studies litigation and settlement (Gabuthy et al., 2021; Massenot et al., 2021) or settlement only (Coursey and Stanley, 1988; Main and Park, 2000, 2002; Inglis et al., 2005). Eisenkopf et al. (2019) account for the impact of negative emotions such as anger, yet not inherent spiteful preferences, and focus on the American fee-shifting rule only. We find that spiteful preferences are consistently associated with higher litigation expenditures and settlement requests. The increase in litigation expenditures is particularly pronounced under the American rule, making the two rules similar for more spiteful players. Additionally, by combining both litigation and pre-trial negotiations, we can compare the expected costs of spitefulness between the two rules. Whereas the harm done to others is similar under both rules, the English rule can protect spiteful players from reducing their own payoffs, as their behavior is less influenced by their spiteful preferences. However, the English rule comes at the cost of overall higher litigation expenditures, driven by less spiteful players. Which fee-shifting rule to implement, therefore, depends both on the distribution of spite and merit in the relevant population.

Third, we also contribute to the recent theoretical literature on spiteful preferences in litigation and settlement. Most importantly, we add to the theoretical work of Chen and Rodrigues-Neto (2023), who show that the effect of negative relational emotions on litigation expenditures depends on the merit of the case and the rule. We also contribute to the theoretical works of Guha, who studies the effects of malice (i.e., spiteful preferences) on

[9] We acknowledge that the experimental work on litigation of Gabuthy et al. (2021) exogenously varies the merit of the case, however, it enters the performance function as an additive constant component only. Consequently, the merit of a case does not influence the marginal effectiveness of effort and thus does not change equilibrium choices.

litigation Guha (2016) or on pre-trial bargaining under the shadow of litigation, however with an exogenous (and unaffected by malice) litigation outcome (Guha, 2019). We provide evidence for some of the theoretical effects of spiteful preferences on litigation and settlement behavior. Additionally, we extend the theoretical literature by studying the effects of spite on pre-trial bargaining (Nash-Demand game), when both the settlement outcomes *and* disagreement outcomes (i.e., expected litigation payoffs) are shaped by spiteful preferences, the merit of the case, and the fee-shifting rule.

The remainder of the paper is structured as follows: In Section 2.2, we briefly summarize the relevant literature. Section 2.3 presents the model. In Section 2.4, we explain the design of the experiment. Section 2.5 shows the experimental results, and in Section 2.6, we conclude.

2.2 Literature

This current paper is related to several strands of literature. In particular, it relates to the law and economics literature on litigation and settlement, as well as to the literature on social and in particular spiteful preferences.

2.2.1 Litigation Literature

A central topic in the theoretical law and economics literature is to model litigation and compare different legal systems. One such way is to model litigation as an all-pay auction (e.g., Baye et al., 2005, 2012), where those who present the best arguments win the dispute with certainty. Another way is to model the process as a Tullock contest (e.g., Plott, 1987; Hirshleifer and Osborne, 2001; Choi and Sanchirico, 2004; Parisi, 2002), where the best arguments win with a certain probability. Arguments are typically modeled as a function of efforts, which represent investments in time, lawyers, and other judicial resources. An important feature of a legal system is the fee-shifting rule, which determines, who has to bear the costs of litigation, and has been studied extensively (e.g., Braeutigam et al., 1984; Chen and Wang, 2007; Baye et al., 2012; Carbonara et al., 2015). Beyond litigation, the theoretical literature has also intensively studied models of settlement (e.g., Schweizer, 1989; Spier, 1992, 1994), including the comparison of the fee-shifting rule (e.g., Reinganum and Wilde, 1986; Hause, 1989). For informative overviews of the litigation and settlement literature, see Spier (2007) and Katz and Sanchirico (2010).

The overall findings concerning fee-shifting are threefold (see Spier, 2007, pp.300-303). First, under the English compared to the American rule, plaintiffs with low-merit cases are less likely to file a lawsuit, while plaintiffs with high-merit cases are more likely to file.

Second, under the English rule, legal expenditures are higher as the marginal benefits have increased and the marginal costs have decreased compared to the American rule. Third, under the English rule, litigation rates are higher.

Several papers analyze public data to evaluate these theoretical predictions empirically. For example, Snyder and Hughes (1990) and Hughes and Snyder (1995) used a legislation change in Florida to study the effect of the fee-shifting rule on the plaintiff's probability to win, jury awards, and out-of-court settlements.[10] Similarly, Fenn et al. (2017) and Helmers et al. (2021) studied litigation expenditures in England and Wales after fee-shifting reforms in 2000 and 2010. Overall, this literature finds that the English rule increases plaintiff's success rates, average jury awards, out-of-court settlements, and average litigation expenditures. The authors argue that these results indicate that the English rule successfully deters low-merit cases from being filed. Our experimental study contributes to this discussion by not only exogenously varying the fee-shifting rule and merit of the case, but also by shutting down the selection effect,[11] as we elicit litigation expenditures independent of participant's settlement outcomes.

In addition to the sparse number of empirical papers, there have been a few experimental approaches studying fee-shifting. For example, Dechenaux and Mancini (2008) conducted an experimental test of the all-pay auction model of litigation by Baye et al. (2005), while Gabuthy et al. (2021) and Massenot et al. (2021) experimentally compared the English and the American rule in a Tullock model. Main and Park (2000) and also Massenot et al. (2021) investigated pretrial bargaining under the American and English rules.[12] Overall, the experimental literature finds that legal expenditures are higher under the English rule, especially for high-merit cases. The evidence for the proportion of cases filed for litigation, however, is mixed. Dechenaux and Mancini (2008) find that under the English rule, fewer cases go to trial, while Massenot et al. (2021) do not find any difference between the two fee-shifting rules. Gabuthy et al. (2021) even find a higher proportion of filed suits under the English rule, especially so for low-merit cases. Unlike Massenot et al. (2021) and Dechenaux and Mancini (2008), we also exogenously vary the merit of the case. Unlike Gabuthy et al. (2021), we vary the merit of the case not as an additive component to the performance function but as an interaction with effort levels, making the merit of the case decisive for

[10] In most of the US, the American rule is used. Florida, however, adopted from 1980 until 1985 the English rule for medical malpractice cases.

[11] Helland and Yoon (2017), for instance, argue that selection effects play an important role in the results of Hughes and Snyder (1995). After correcting for selection effects they can only reconfirm that the English rule increases out-of-court settlements but not its impact on trial awards and litigation expenditures.

[12] Other papers study the 50 percent rule in the lab (Thomas, 1995), pretrial bargaining with a shadow of the future (Coursey and Stanley, 1988; Main and Park, 2000), and negotiations and conflict under the shadow of the future (Main and Park, 2002; McBride and Skaperdas, 2014; McBride et al., 2017).

equilibrium choices. Unlike the existing experimental work, we also study the interaction of the fee-shifting rule and merit with spiteful preferences both for litigation expenditures and settlement requests.

Notably, many of the theoretical papers assume plaintiffs and defendants to be self-interested and without any biases or social preferences.[13] These assumptions, however, are strongly contrary to the findings in experimental economics as outlined below.

2.2.2 Literature on Spiteful Preferences

Extensive experimental literature has provided evidence that subjects do not merely selfishly maximize their own payoffs but also care about others' payoffs and thus exhibit social preferences (for an overview, see Cooper and Kagel, 2016). Not only positive (Andreoni, 1989) but also negative social preferences have been shown to influence behavior. For example, Andreoni et al. (2007), Cooper and Fang (2008), Herrmann and Orzen (2008), Kimbrough and Reiss (2012), Bartling et al. (2017), and Kirchkamp and Mill (2021a) used experiments to show that subjects have spiteful preferences and that these lead to more competitive behavior. Similarly, Abbink and Sadrieh (2009), Abbink and Herrmann (2011), and Bauer et al. (2023) show in experiments that subjects display nasty and antisocial behavior. A key insight from this literature is that spiteful preferences influence behavior in many economic settings.

Consequently, the theoretical literature started to incorporate the influence of spiteful preferences in litigation (Guha, 2016; Chen and Rodrigues-Neto, 2023), standard bargaining (Montero, 2008; Guha, 2018), and pre-trial bargaining settings (Guha, 2019).[14] For instance, Guha (2016) studies the effect of malicious preferences on litigation behavior. She develops her own model of litigation and models malicious preferences as additional utility coming from the payment endured by the defendant. In a later study, Guha (2019) incorporates spitefulness in dynamic pretrial settlements under the threat of litigation. Here, she introduces malice as utility coming from the opponent's litigation costs *and* costs of waiting for a resolution. Additionally, the utility of the disagreement outcome is influenced by the degree of malice. The outcome itself, however, is exogenous and unaffected by malice. Unlike Guha (2016) and Guha (2019), we use a rather standard model of litigation (Hirshleifer and Osborne, 2001) and bargaining (i.e., a Nash demand game), vary the merit of a case, and study not only the American but also the English fee-shifting rule. Different from the models

[13] For an exception see Heyes et al. (2004), who assume agents to be risk-averse, Baumann and Friehe (2012) model agents to have emotions, and Guha (2016, 2019) and Chen and Rodrigues-Neto (2023) who assume malicious agents.

[14] In auction settings, Morgan et al. (2003b), Mill (2017b), Bartling et al. (2017), and Kirchkamp and Mill (2021a) used theoretical means to show that spiteful preferences lead to overbidding.

of malicious preferences of Guha (2016) and Guha (2019), our model of spiteful preferences focuses on the final payoffs and not only on the costs endured. Importantly, unlike Guha (2019), we model settlement with an endogenous disagreement outcome – the expected litigation outcome – which is determined in equilibrium and shaped by the players' level of spite, the merit of the case, and the fee-shifting rule.

Chen and Rodrigues-Neto (2023) study the interaction of emotions and the fee-shifting rule in litigation settings. Litigants obtain additional emotion-based utility depending on the final payoff of the opponent, which can be either positive or negative. They define a generic model that captures, among several others, the Tullock contest success function.[15] They find that negative emotions amplify the costs of fee-shifting – this implies that the increase in litigation costs due to negative emotions is higher under the English rule compared to the American. Unlike Chen and Rodrigues-Neto (2023), we also study spiteful preferences and the interaction with the fee-shifting rule for pretrial bargaining under the threat of (spiteful) litigation.

Finally, this paper is also related to the theoretical contest literature with spillovers (Chowdhury and Sheremeta, 2011a,b; Baye et al., 2012; Betto and Thomas, 2024). When opponents are motivated by spiteful preferences, contestants' own choices create spillovers for their opponent's utility, and vice versa. Thus, we provide experimental evidence that spite-driven spillovers influence behavior in litigation settings.

We are not aware of any experimental study, which investigates the impact of spiteful preferences in a litigation or settlement setting. Most closely related is the study of Eisenkopf et al. (2019), which focuses on the impact of emotions in a litigation setting. They do not find any effect of emotions on litigation expenditures. The authors induce emotions through a pre-litigation stage, where players can steal money from their opponent. Instead of studying the impact of emotions, we focus on inherent antisocial preferences. Furthermore, unlike relying on non-incentivized self-reports of emotions, we primarily rely on an incentivized behavioral spite task. Lastly, we also investigate the interaction of spite with the fee-shifting rule and the merit of the case.

Overall, we add to the literature by providing a thorough (experimental) investigation of how spitefulness interacts with the American and English rule with an exogenous variation in the merit of the case. We do this both for litigation litigation behavior and pre-trail bargaining under the shadow of endogenous spiteful litigation.

[15]To ensure an interior solution under the English rule, the authors assume the exponent of the CSF to be smaller than 1. We, instead, rely on the most commonly used Tullock model, the lottery contest, where the exponent equals 1. We ensure a (corner) solution by constraining the litigation expenditures upwards to the winning prize.

2.3 Theoretical Models

In this section, we build a theoretical model to derive predictions in order to guide the interpretation of our experimental results. Our aim is not to present an all-encompassing model of litigation but rather to provide some intuition for what could be expected in our experiment. Despite this modest aim, the model generally provides valuable insights into litigation and settlement behavior when agents have spiteful preferences.

We model both litigation and settlement behavior under the American and English fee-shifting rules and vary the merit of the case. We incorporate spiteful preferences both for litigation expenditures (in Section 2.3.1) and settlement requests (in Section 2.3.2).

2.3.1 Litigation Model

To model litigation, we use a model similar to Hirshleifer and Osborne (2001). To model spiteful preferences, we build on Morgan et al. (2003b).[16]

We assume two litigants, i and j, who denote the defendant and the plaintiff, respectively. Both litigants make a decision upon their effort for litigation $e_k \in [0, \overline{e}]$ with $k \in \{i, j\}$. The litigation effort represents the cumulative effort invested in the litigation process and aims to reflect the quality of the argument brought forward in court. The litigation effort includes – among other things – the personal effort in finding and providing evidence, the cost for the lawyer, and the time invested in making the arguments.

Both i and j litigate for a prize of common value $W \in \mathbb{R}$. We further assume risk-neutral and spiteful litigants, who are spending at most the value of the prize, i.e., $\overline{e} = W$.[17]

In court, the judge makes a decision to whom to assign the prize, based on the arguments and also based on the commonly known merit of the plaintiff's case $q \in [0, 1]$. The merit of the case can be interpreted as the general tendency of a particular judge to rule in favor of the plaintiff. It can also be considered as argument weighting due to fairness or it could represent who 'truly' deserves to win the case (in a world with perfect information).

In particular, we use the following contest success function for player i, which is a special case of the contest success function suggested by Hirshleifer and Osborne (2001):

$$p_i(e_i, e_j, q) \colon = \frac{(1-q) \cdot e_i}{q \cdot e_j + (1-q) \cdot e_i}$$

[16]See also Bartling et al. (2017), Mill (2017b), Mill and Morgan (2018), and Kirchkamp and Mill (2021a) for the use of this model.

[17]Introducing this upper bound guarantees the existence of a Nash equilibrium under the English fee-shifting rule. Constraining the litigant's expenditures also reflects reality in that they cannot spend infinite resources.

and correspondingly the probability of player j to win the argument is denoted by $1 - p_i(e_i, e_j, q)$. If both players do not invest anything, i.e., when $e_i = e_j = 0$, then winning chances are simply determined by the merit of a case: $p_i(0,0,q) = 1 - q$ and $p_j(0,0,q) = q$. Several aspects of this simple contest success function are worth pointing out:

- If either one of both players drops out of litigation (i.e., $e_k = 0$), the probability to win will be 1 for the other player.
- If both players provide equally good arguments (i.e., $e_i = e_j$), the probability to win for player j depends solely on the merit of the plaintiff's case q (and $1 - q$ for player i).
- If the merit of the plaintiff's case is zero (i.e., there is absolutely no merit to the case), player i (the defendant) wins with certainty.
- Correspondingly, if the merit of the defendant's case is zero (i.e., the judge is purely in favor of the plaintiff), player j wins with certainty.

After the judge's ruling, the winner obtains the prize W and the loser does not. Under the American rule, the winner and loser each have to pay their effort costs. Under the English rule, the loser has to pay his own costs and compensate for the entire effort costs of the winner.

Moreover, we assume that agents exhibit external preferences, i.e., their utility is influenced by the payoff of the other litigant. We use a model suggested by Morgan et al. (2003b), where agents receive additional disutility from the opponent's payoff and hence additional utility from the opponent's negative payoff (i.e., costs).[18] We define $\alpha \in (0,1)$ and the opponent's payoff ϕ, which results in the defendant's additional utility $\nu_i(\alpha_i, \phi_j) = -\alpha_i \cdot \phi_j$ and in the plaintiff's additional utility $\nu_j(\alpha_j, \phi_i) = -\alpha_j \cdot \phi_i$. For simplicity, we assume that $\alpha_i = \alpha_j = \alpha$. Hence, the overall utility (u_i) of litigation of agent i can be written as:

$$u_i(e_i, e_j, q, \alpha) = \underbrace{p_i(e_i, e_j, q)\cdot}_{\text{Probability of winning}} \underbrace{\left(\underbrace{W - 1_{American} e_i}_{\text{Payoff}} \underbrace{+\alpha \cdot (e_j + 1_{English} e_i)}_{\text{Utility due to spite}} \right)}_{\text{Utility in case of winning the case}} \tag{2.1}$$

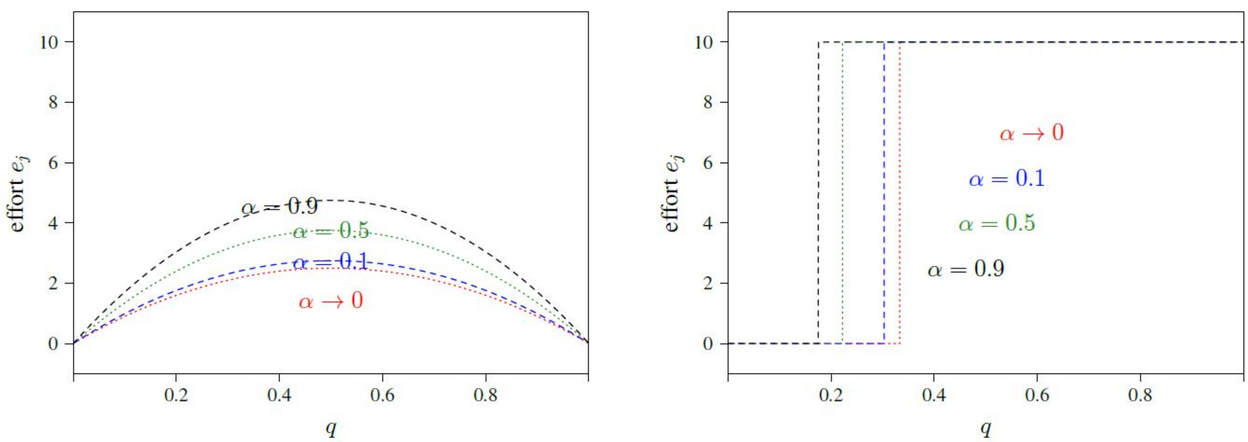
$$+ \underbrace{(1 - p_i(e_i, e_j, q))\cdot}_{\text{Probability of losing}} \underbrace{\left(\underbrace{-e_i - 1_{English} e_j}_{\text{Payoff}} \underbrace{-\alpha \cdot (W - 1_{American} e_j)}_{\text{Disutility due to spite}} \right)}_{\text{Utility in case of losing the case}}$$

[18] Note that this definition of the spite motive builds on the absolute payoff of the opponent and not on the payoff differences. Hence, it is distinct from disutility coming from inequality aversion.

Similarly, the utility (u_j) of litigation of agent j is $u_j(e_j, e_i, q, \alpha) = u_i(e_j, e_i, 1-q, \alpha)$. We assume that both litigants simultaneously maximize their utility and simultaneously decide on their litigation expenditures conditional on their opponent's best response (Nash equilibrium).[19] In the following, we refer to the plaintiff (player j). Hence, we speak of low merit when q is small and of high merit when q is high.

Figure 2.1 depicts the static symmetric equilibrium expenditures under the American or English fee-shifting rule for player j. Formal propositions and their derivations can be found in Appendix B.1.1. In the following, we describe the equilibrium expenditures and derive the theoretical predictions (see Appendix B.2.5 for the formal derivations, where necessary).

Figure 2.1: Equilibrium litigation expenditures for spiteful litigants



Note: Equilibrium litigation expenditures (e_j) under the American (left) and English (right) fee-shifting rule with $W = 10$ for different merits q and different spite levels α (see Proposition (1) and (2) in Appendix B.1.1). Note that the vertical lines in the right panel are presented just for illustration purposes (i.e., 0 and 10 are optimal but not the values in between).

Under the American rule, players have to bear their own litigation costs, independent of the litigation outcome. Therefore, litigation expenditures are highest when none of the players has a relative advantage, i.e., when both players have the same merit $q = 0.5$. That is because for low-merit cases ($q < 0.5$), there is a smaller chance to win and thus it does not pay off to spend much on litigation. Moreover, for high-merit cases ($q > 0.5$), winning probabilities are already high – also because the opponent does not invest much; thus, fewer own expenditures are needed.

Under the English rule, players do not have to bear their own expenditures in case of winning, and thus the expected litigation expenditures are low if winning chances are high.

[19]The utility function (including the spillover parameters) under the American rule satisfies the conditions Chowdhury and Sheremeta (2011b) lay out for the existence of a unique symmetric equilibrium. We show the existence of a unique equilibrium under the English rule in Appendix B.2.2.

Yet, if winning chances are low, expected costs are high because the loser additionally has to bear the winner's expenditures. Therefore, if the prospects of winning are good enough, players fully invest in litigation. However, if the prospects of winning are too low, players are incentivized to at least save their own litigation expenditures, and reduce their spendings to zero.[20]

The benefit of winning is higher under the English rule compared to the American since the loser has to pay all the costs. Therefore, overall, agents spend more resources under the English compared to the American fee-shifting rule:

Hypothesis 1.1. *Average litigation expenditures of all merit levels q are higher under the English fee-shifting rule than under the American fee-shifting rule.*

Being spiteful introduces additional disutility from losing and utility gains from winning. Hence, spite widens the prize gap, and thus litigation expenditures increase. Under the American rule, expenditures increase proportionally depending on the merit. Under the English rule, spite influences expenditures only for cases, where agents have not fully invested yet, shifting this threshold to ever lower merit cases (see Figure 2.1). Since agents have not fully invested only for low-merit cases, spite exclusively affects a specific range of low-merit cases under the English rule. Aggregated over all merit levels q, average litigation expenditures are higher for more spiteful agents.[21]

Hypothesis 1.2. *Under the American fee-shifting rule, average litigation expenditures are higher for more spiteful agents at every merit level.*

Hypothesis 1.3. *Under the English fee-shifting rule, average litigation expenditures over all merit levels q are higher for more spiteful agents. This increase is driven by an increase at a specific range of low-merit levels only while there is no increase at high-merit levels.*

Next, we study whether this increase in litigation expenditures for more spiteful agents is more pronounced under the American or English fee-shifting rule. To get a benchmark prediction, we compare the average expenditure of a fully spiteful agent ($\alpha \to 1$) to the average expenditure of a fully non-spiteful agent ($\alpha \to 0$) over all merit levels q. In this specific case, there is no difference in the average increase for being spiteful compared to not being spiteful between the two rules.

Hypothesis 1.4. *There is no difference in the increase of the average litigation expenditure over all q between the English and American fee-shifting rule for a non-spiteful ($\alpha \to 0$) compared to a fully spiteful ($\alpha \to 1$) agent.*

[20] We acknowledge that this bang-bang property is an extreme feature that our simplified model produces. In reality, behavioral shifts are likely more gradual, influenced by factors such as the way of winning or losing, which may create additional internal utilities or even change the verdict (i.e., W).

[21] A more detailed explanation of the incentives under the English rule can be found in Appendix B.1.1.

2.3.2 Settlement Model

In some cases, agents may not want to litigate.[22] To avoid litigation, agents can also settle the dispute. We model settlement behavior as a standard Nash-Demand game: Two agents make a suggestion of how to split a good W by requesting a certain amount of this good (s_k with $k \in \{i, j\}$). If both the requests of i and j sum to W, i.e., $s_i + s_j = W$, the requests are granted. If both the requests are in sum less than W, i.e., $s_i + s_j < W$, both obtain their request plus half of the leftover as their payoff, i.e., $\Phi_i = s_i + \frac{W - s_i - s_j}{2}$. If, however, the sum of both the requests exceeds W, i.e., $s_i + s_j > W$, then no settlement is reached, and agents have to litigate for W.

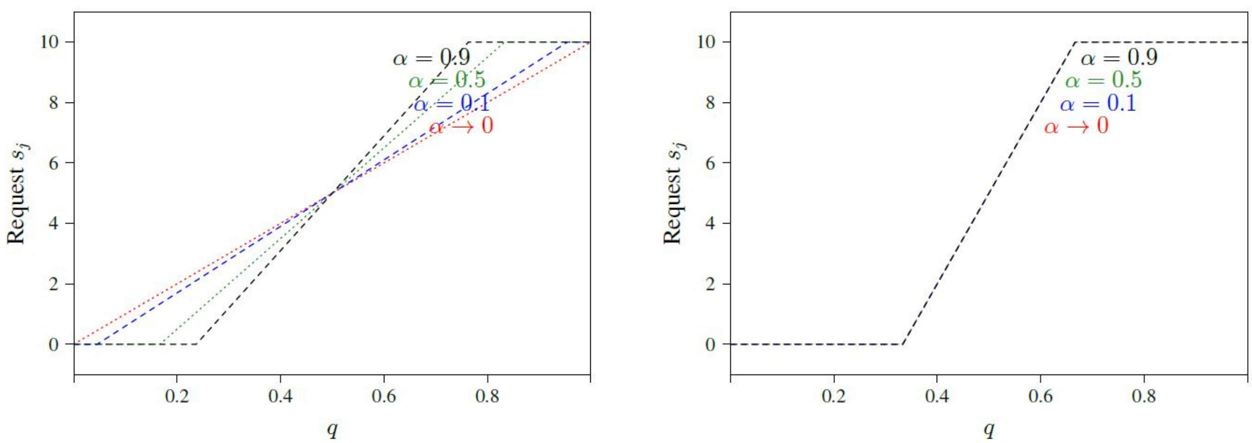
Spiteful preferences affect the utility function both in case the settlement is successful and when it is not. When settlement is successful, player i receives utility from her share s_i of the pie. Additionally, due to spite, she receives disutility from player j's share of the pie. This settlement utility is described by $u_i^{settlement}(s_i, s_j, \alpha) = s_i - \alpha s_j$ and for player j by $u_j^{settlement}(s_i, s_j, \alpha) = s_j - \alpha s_i$. If the settlement is not successful, agents have to litigate for W. The expected payoffs from the litigation stage are called the disagreement values $d_i(e_i^*, e_j^*, q, \alpha)$ and $d_j(e_i^*, e_j^*, q, \alpha)$. Note that both the disagreement value and the settlement utility are directly affected by the spite parameter α.

We rely on the Nash-Demand solution (Nash, 1950), and more specifically, on the efficient pure-strategy Nash equilibrium, where we maximize the function $f = (u_i^{settlement} - d_i)(u_j^{settlement} - d_j)$ under the constraints that $s_i + s_j = W$ and $s_i, s_j \in (0, W)$.[23] Figure 2.2 shows the static symmetric equilibrium settlement requests for player j under the American and English fee-shifting rules. In the following, we describe the equilibrium settlement requests and derive the theoretical predictions (see Appendix B.2.6 for the formal derivations, where necessary).

[22] For instance, under the English fee-shifting rule, litigation leads to negative payoffs for not high enough merit levels for one of the parties.

[23] While this approach does not yield any settlement failures, it nonetheless provides useful comparative statics with respect to spiteful preferences and their interaction with the fee-shifting rule and merit.

Figure 2.2: Equilibrium settlement requests for spiteful litigants



Note: Equilibrium settlement requests (s_j) under the American (left) and English (right) fee-shifting rule with $W = 10$ for different merits q and different spite levels α (see Proposition 3 and 4 in Appendix B.1.2).

In the Nash-Demand Game, settlement requests are determined by the players' bargaining power. Under both rules, bargaining power is shaped by the expected litigation outcomes and therefore directly by the merit of the case. Consequently, requested settlement amounts increase in the merit. Whenever the expected payoff of litigation is negative, the requested amounts are zero, where the opponent would then request the full amounts. In the efficient pure-strategy Nash equilibrium, all resources are allocated without waste among the two players. Therefore, there are no differences in average settlement requests over all merits q under the American and the English fee-shifting rule.

Hypothesis 2.1. *There is no difference in average settlement requests over all merits q between the American and the English fee-shifting rule.*

When players have spiteful preferences, both their expected utility from litigation – and hence their expected disagreement outcomes – and their utility from the settlement outcomes are influenced. In the litigation stage, spite further widens the gap between the winning and losing utility and hence the changes in the expected litigation utilities depend on the winning probabilities. Winning probabilities are directly shaped by the merit of the case (and in the symmetric equilibrium unaffected by the players' level of spite). Consequently, for low-merit cases ($q < 0.5$), spite decreases the expected utility even further, whereas for high-merit cases ($q > 0.5$), spite increases the expected utility of litigation. Therefore, more spiteful players demand more for high-merit cases and less for low-merit cases, which is reinforced by the opponent's requests, who has the 'opposite' merit $1 - q$.

In the settlement stage, any concession to the opponent reduces the own utility due to spite. The player who allows the opponent to have a larger part of the good W receives more disutility (in the symmetric spite case). Thus, for low-merit cases ($q < 0.5$), the disutility is higher for the disadvantaged player and consequently, spite increases the requested amount. For high-merit cases ($q > 0.5$), the player anticipates that a spiteful opponent will not want to give her much, and thus spite decreases her demand. That works as a direct attenuating effect to the effect of spite on the expected litigation utility. Under the American rule, the effect of spite on the disagreement outcome in the litigation stage prevails, while under the English rule, these effects are perfectly attenuated by spite's effects in the settlement stage.

Hypothesis 2.2. *Under the American fee-shifting rule, the average settlement requests for low merits ($q < 0.5$) are lower for more spiteful agents, while for high merits ($q > 0.5$) they are higher.*

Hypothesis 2.3. *Settlement requests under the English fee-shifting rule are the same for more spiteful and less spiteful agents.*

Finally, there is no difference in the average influence of spite on the settlement requests over all merits q, because all resources are always allocated without waste in the efficient Nash-Demand game solution.

Hypothesis 2.4. *There is no difference in the change of average settlement requests over all merits q between the English and American rule for non-spiteful compared to spiteful agents.*

2.4 Experiment

In this section, we describe the design of the litigation experiment (in Section 2.4.1), our measures of spiteful preferences (in Section 2.4.2), subject recruitment (in Section 2.4.3), payment (in Section 2.4.4), and the procedure of the experiment (in Section 2.4.5).

2.4.1 Litigation Experiment

To test the theoretical predictions, we manipulated the fee-shifting rule, which was either *American* or *English*, as well as the merit of the case. The fee-shifting factor was implemented in a within-subjects design, i.e., every subject made all decisions both for the American and the English fee-shifting rule. To cope with order effects, we counterbalanced the order of the fee-shifting rule: Half the participants made decisions under the American regime first and then under the English one whereas the other half of the participants made decisions under the English regime first and then under the American one.[24]

To have a clear design and to exclude effects of winning or losing (e.g., hedging effects or retaliative motives), the experiment was conducted as a one-shot game. This means that

[24]Appendix B.4.1 provides evidence of the absence of an order effect.

subjects made all their decisions only once and that there was no feedback between any of the decisions.

In addition to the litigation decisions, we also elicited settlement behavior.[25] Thus, subjects had to make two decisions: the litigation and the settlement decision. The litigation stage was played under each regime first, and only then subjects were instructed and asked to make the decision for the settlement stage. This has three advantages: 1) it ensures that subjects do indeed follow backward induction, 2) it ensures that litigation behavior is not impacted by the mere failure of the settlement stage, i.e., subjects are not driven by anger due to a failed settlement[26] and, more importantly, 3) it ensures the experiment not to have a selection bias – i.e., all subjects litigate and not only those who fail settlement. So, subjects made a litigation decision first and then they were asked to settle the dispute under the shadow of litigation –i.e., if the settlement stage was payoff-relevant and they settled successfully, this settlement represented their payoff. However, if they failed to settle, the outcome of the litigation stage would be payoff-relevant. No information regarding the other players' choices was provided between the two stages.[27] Thus, all observations are statistically independent.

The settlement was designed as a standard two-player Nash-Demand game as described in the model Section 2.3.2. The litigation stage was played as a standard two-player Tullock contest. To ensure that subjects do not end up with a negative payoff, they were always endowed with 10 tokens. In addition, subjects competed in the litigation stage for a prize of 10 tokens, and no subject was endowed with the litigated object to reduce biases due to loss-aversion (Kahneman and Tversky, 1984; Tversky and Kahneman, 1992), an endowment effect (Kahneman et al., 1990; Plott and Zeiler, 2007), and more generally reference-dependent preferences (Kőszegi and Rabin, 2006).

Furthermore, all subjects had to make five decisions in each stage – settlement and litigation – under each regime – English and American. The decisions differed only by the parameter q – representing the merit of the case from the plaintiff's point of view – where low merit corresponds to a low q and high merit to a high q. The five chosen levels of q were $q \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$. To cope with order effects, the order of the presented qs was randomized by subject. Figures B.8 and B.9 in Appendix B.6 show the interface for the litigation and settlement decision under the English rule for q of .5, respectively. As

[25] To ensure incentive compatibility, we randomly only paid either the resulting payoffs of the litigation or the settlement stage. For a more detailed description of the payment procedure, see Section 2.4.4.

[26] In this way we can isolate the effect of inherent spiteful preferences.

[27] By playing litigation first, but not receiving information about their opponent's behavior, players are aware and familiar about the litigation stage but do not know the exact outcome. Instead, as in a real world setting, they have to rely on their beliefs about their opponent's litigation behavior when deciding on their settlement requests.

subjects did not get any feedback between the decisions – in fact, subjects were informed about the outcome of all tasks only after a day – the decisions represent a strategy method approach (Selten, 1967). Overall, subjects made 2 (Regime: English, American) x 2 (Stage: Settlement, Litigation) x 5 (Merit $q \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$) = 20 decisions.[28] Hence, all of our treatment manipulations concerning the merit and the fee-shifting rule are within-subjects.

To reduce experimenter demand effects, we instructed subjects on an abstract level, i.e., we did not use words like litigation, settlement, court, American, English, plaintiff, defendant, etc. Instead, the litigation stage was presented as 'Task A,' and the settlement stage was presented as 'Task B'. Subjects were instructed in the litigation stage as typically done in contest experiments, and in the settlement stage, they were instructed as usually done in Nash-Demand experiments (see also the instructions in Appendix B.7.1).

2.4.2 Spiteful Preferences Measures

After the litigation experiment, we elicited spiteful preferences via two different methods.[29] Specifically, we used the *Spite-Task* (Mill and Morgan, 2021; Kirchkamp and Mill, 2021a) and the *Spite-Questionnaire* (Marcus et al., 2014). Additionally, we employed the *SVO-Task* (Murphy et al., 2011a; Murphy and Ackerman, 2014) to elicit prosocial preferences, which we use as a robustness check for the effects of spiteful preferences.

Spite-Task: We use the Spite-Task (Mill and Morgan, 2021) to measure spiteful preferences towards the opponent in the experiment, which is similar to the SVO-Slider measure (Murphy et al., 2011a). In the Spite-Task (see Table 2.1), subjects make three money distribution decisions. While the allocation that maximizes their opponent's payoff also maximizes their own, subjects can intentionally reduce their opponent's payoff. Depending on the allocation decision, this reduction is either costless or comes with a personal cost. Therefore, when subjects choose to reduce their opponent's payoff, they do so because they actively want to harm the other player. Consequently, we interpret any deviation from the payoff-maximizing allocations as spitefulness. We made participants aware that it was randomly

[28] Our setting encourages cold decision-making rather than hot decision-making, which may make it more difficult to identify emotion-based spite effects. However, we argue that we can still identify inherent preference-based spitefulness in our setting. Further, the cold-decision making makes our design cleaner as other factors associated with hot decision-making could have confounded our results (such as other emotion-based social preferences).

[29] Additionally, we aimed to manipulate the extent of spite by excluding social preferences altogether. For this purpose, participants were either matched with a computer or another human participant. The manipulation, however, seems not to have worked as the manipulation was too weak. We present the results of the manipulation in Appendix B.5 and provide a detailed discussion of why we believe the manipulation failed.

determined whether their own or their opponent's allocation decision would be implemented.

Table 2.1: Spite measure

You receive	70	70	70	70	70	70	70	70	70
	○	○	○	○	○	○	○	○	○
Other receives	100	98	96	94	92	91	89	87	85
You receive	70	68	65	62	60	58	55	52	50
	○	○	○	○	○	○	○	○	○
Other receives	100	96	92	89	85	81	78	74	70
You receive	100	100	100	100	100	100	100	100	100
	○	○	○	○	○	○	○	○	○
Other receives	100	98	96	94	92	91	89	87	85

Note: The table depicts the three allocation decisions in the Spite-Task, where the players decide among nine possible allocations in each. The upper rows show their own payoff for the deciding player, while the bottom rows show their opponent's payoff.

In the Spite-Task, the *spite score* indicates how much the player reduced the payoff of their opponent relative to the maximum possible amount. Players can reduce their opponent's payoff from 0 and 60 points in all three decisions combined and, therefore, the spite score ranges from 0 to 1.

Spite-Questionnaire: The additional measure of spitefulness is a questionnaire by Marcus et al. (2014), where participants are asked to rate 17 statements. Here are two examples:[30]

- I would be willing to take a punch if it meant that someone I did not like would receive two punches.
- I would be willing to pay more for some goods and services if other people I did not like had to pay even more.

Participants were asked to indicate their agreement on a scale between 1 and 5. Higher scores on the scale indicate more spitefulness. This task's measure of spitefulness is the average agreement with the statements.

SVO-Task: To measure prosocial preferences, we used the 6-items primary scale of the SVO Slider Task (Murphy et al., 2011a; Murphy and Ackerman, 2014). The primary scale of the SVO-task consists of six distribution-decisions among nine possible allocations. Based on these answers, a continuous variable is calculated (i.e., the SVO-angle). This variable represents a participant's prosocial preference and ranges from -16.26 to 61.39, where a higher value represents more prosocialness.

[30] All questions are shown in Appendix B.7.4.

2.4.3 Subject Recruitment and Selection

The experiment was conducted online and subjects were recruited via Amazon's Mechanical Turk (MTurk).[31] We use an MTurk sample because they are typically more diverse in terms of age, ethnicity, education, and geographical location, and therefore tend to better represent the US population than usual student samples (Buhrmester et al., 2011; Berinsky et al., 2012; Paolacci et al., 2010). Several studies show that the data obtained in MTurk is similar to data typically obtained in laboratory experiments (Paolacci et al., 2010; Buhrmester et al., 2011; Horton et al., 2011; Berinsky et al., 2012; Arechar et al., 2018a).

An additional advantage of employing an online design is that we can sufficiently ensure participants' anonymity, as we only have access to their MTurk-ID. This anonymity might enhance the reliability of results regarding subjects' litigation and settlement behavior and especially regarding their spiteful preferences. Furthermore, we minimize reciprocity concerns because participants do not meet each other in the online context, and we do not communicate the identity of their matched partner. Finally, through this anonymity, we can also exclude social ties and peer effects.

One obvious disadvantage of such an online setting is that subjects might pay less attention. To tackle this potential issue, and to ensure a high-quality sample, we restrict recruitment to US-based individuals, which have an approval rate of at least 97% and more than 500 approved HITs.[32] Additionally, subjects had to answer incentivized control questions after reading the instructions.[33]

2.4.4 Payment

To ensure that all decisions are incentive-compatible and equally relevant, we paid out one randomly picked decision.[34] Subjects were told that only one scenario (q) of one stage – i.e., either the litigation or the settlement stage – under one fee-shifting rule would be randomly picked for payment. The matching of players was randomly performed after all decisions were made. As all subjects had to indicate their decisions for all scenarios (q), we assigned

[31]The platform is frequently used by economists (e.g., DellaVigna and Pope, 2018; Horton et al., 2011) and other social scientists (e.g., Jordan et al., 2016b, 2017; Peysakhovich et al., 2014; Rand et al., 2014; Suri and Watts, 2011; Mao et al., 2017).

[32]Subjects' location is verified through their IP addresses. Requesters can review the work done by MTurkers and decide to approve or reject the work. Approved work is paid as indicated in the contract, and rejected work is not paid. Hence, higher approval rates of workers indicate a higher quality of work.

[33]In a second wave, we further excluded subjects who used a VPN from outside the US, subjects on mobile devices, and bots. Additionally, every participant had to answer the control questions correctly before being able to proceed with the experiment. In the second wave, we also elicited risk aversion (discussed in Appendix B.4.4) and find that it does not interact with the influence of spiteful preferences. The overall results of the two waves are qualitatively comparable (see Appendix B.4.3).

[34]See Azrieli et al. (2018b) for a detailed argument.

each randomly matched pair one q, and each subject was randomly assigned the role of either the plaintiff (i.e., the decision for q was payoff-relevant) or the defendant (i.e., the decision for $1-q$ was payoff-relevant). The payment was executed a day after all subjects had made their decisions.

2.4.5 Procedure

Subjects were recruited for this experiment via Amazon Mechanical Turk and were directed to an external survey link. As soon as subjects arrived at our platform, they were asked for their individual MTurk-ID to ensure payment at the end of the experiment. After giving consent to participating in the experiment, subjects were asked to answer several socio-demographic questions, i.e., age, gender, education, and ethnicity. Thereafter, subjects were instructed with the experimental task and had to answer incentivized control questions (each control question gave an additional 5 dollar cents). After making all decisions of the litigation experiment, subjects were instructed for the SVO-Task and the Spite-Task. They stayed in the same pairs as in the litigation experiment. Some participants also took part in the risk task (which we explain and discuss in Appendix B.4.4). After answering the Spite-Questionnaire, subjects were directed back to Amazon Mechanical Turk. The procedure is depicted in Figure B.10 in the appendix.

2.5 Results

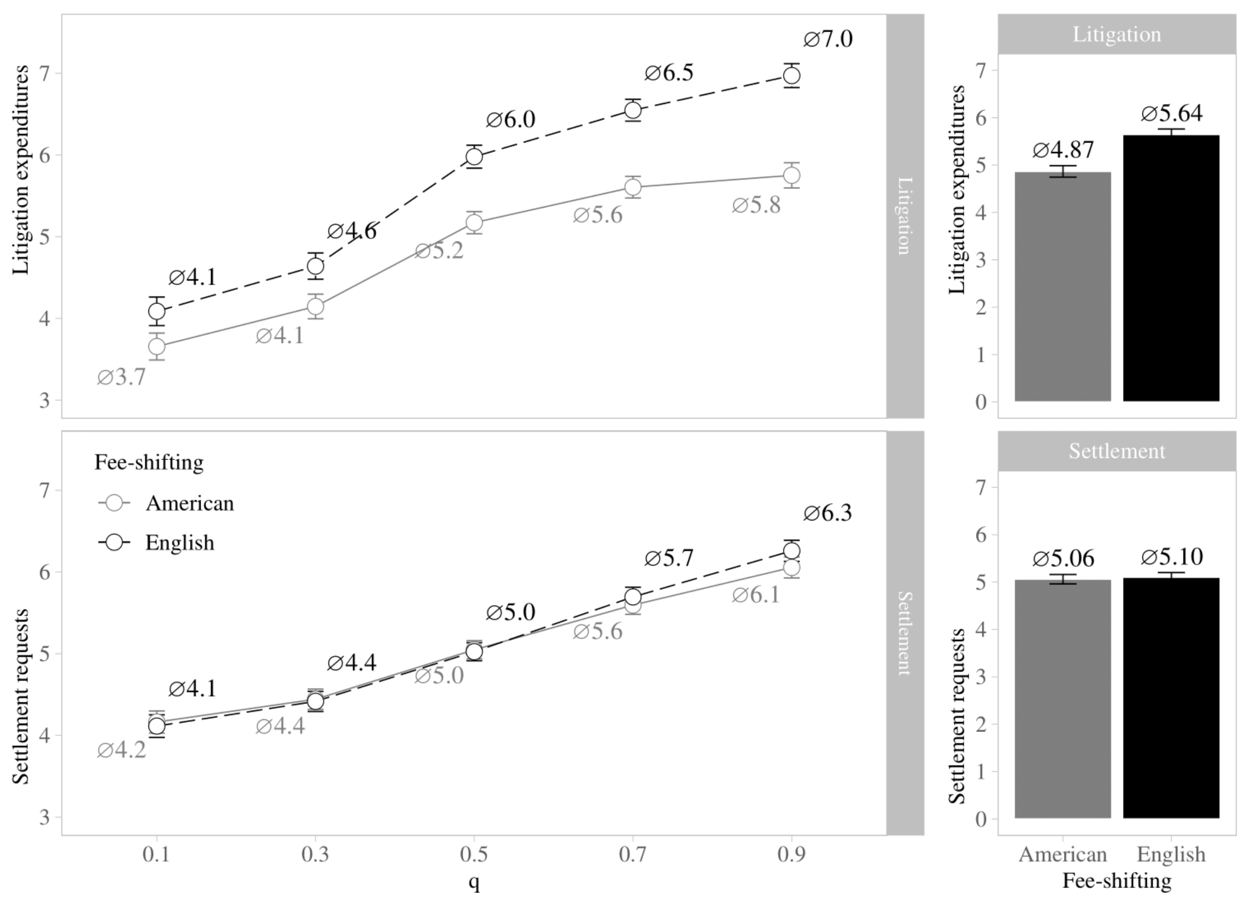
We conducted the experiment in two waves: The first wave took place in November 2017, and the second wave in January 2021. We recruited 1635 participants and the experiment was implemented using the online survey tool Qualtrics. The entire experiment lasted for about 30 minutes (SD = 18.0). Median earnings of participants were \$2.90 (including a show-up fee of \$1) resulting in an average hourly wage of \$7.13, which is more than the median hourly income of a typical MTurker. We had 51% female participants, participants' age ranged from 18 to 81, and 78 % of participants reported to have at least a college degree.

Throughout this entire results section, we present the results based on the plaintiff's view of merit, where low merit corresponds to a low q and high merit to a high q. Since all subjects had to indicate their decision based on this view, we can classify all subjects using the same merit classification. To derive the observed behavior from the defendant's view of merit, q can be swapped with $(1-q)$ in all the results and figures.

2.5.1 American vs. English Fee-shifting

We start with studying differences in litigation and settlement behavior between the American and English fee-shifting rules. The left part of Figure 2.3 shows litigation expenditures and settlement requests for both fee-shifting rules conditional on the merit q, while the right part shows the decisions for both regimes as an average over all possible (uniformly distributed) merits q.

Figure 2.3: Litigation expenditures and settlement requests



Note: The figures to the left depict the litigation expenditures and settlement requests by fee-shifting rule as a function of q, while the figures to the right depict the aggregates. The panels on the top show the litigation effort, while the panels on the bottom illustrate the settlement requests. Grey solid lines depict the behavior under the American fee-shifting rule, while black dashed lines indicate the response under the English fee-shifting rule in each panel. The error bars indicate the 95% confidence intervals.

As a first step, we focus on litigation expenditures. We find that, on average, subjects invest significantly ($t(1634)$= -13.2, $p < 0.001$) more under the English (5.64 tokens) compared to the American regime (4.87 tokens), as Hypothesis 1.1 suggests. Not only on average but for any merit level, subjects significantly invest more under the English rule, including for low-merit cases. That suggests that the English fee-shifting rule does not help in preventing frivolous (low-merit) litigation, as theory would predict. On the contrary, the English rule seems to lead to higher investments in the litigation process for cases with little merit. We report on the regression analysis of the differences between the English and American rule as a function of the merit q in Appendix B.3.1.

Overall, the behavior of subjects in the experiment (see Figure 2.3) does not seem to align with the functional theoretical predictions of the American nor the English rule (see Figure 2.1). For instance, under the English rule, our theory predicts either full or no expenditures, depending on the merits. Subjects in the experiment, however, increase their litigation expenditures gradually, which leads to higher investments for low-merit cases than predicted and less for high-merit cases. One part of this behavioral deviation from the theoretical benchmark can be explained by prospect theory (Tversky and Kahneman, 1992): subjects underestimate their winning chances for high-merit cases and hence are more careful in their litigation expenditures, while they overestimate their winning chances for low-merit cases and thus invest more, even though they likely have to carry their own (and opponent's) costs. This probability distortion may especially matter for the English rule, because the incentives to win are higher, which leads to a higher (perceived) marginal increase in the expected utility of an additional unit spent for low-merit cases compared to the American fee-shifting rule. Additionally, other non-monetary utilities, which depend on the way one wins or loses, may lead to a gradual increase in litigation expenditures rather than an all-or-nothing strategy.

Under the American rule, even though there is an apparent increase in litigation effort from low to medium merit, there is no decrease from medium to high merit. The absence of this decrease might be explained by anticipated regret (Filiz-Ozbay and Ozbay, 2007, 2010; March and Sahm, 2017), where subjects anticipate regret of not having invested more if they could have won.

Next, we compare how settlement requests, on average, differ between the English and American fee-shifting rule. We do not find any significant differences ($t(1634)$= -1.3, $p \geq 0.05$) between subjects' average settlement requests under the American (5.06 tokens) compared to the English regime (5.10 tokens). Once again, we observe this non-difference not just on average, but across all merit levels individually. Thus, the data does not give suggesting evidence to reject Hypothesis 2.1, denoting no difference in the average settlement requests between the American and the English rule.

Concerning the theoretical functional form of settlement requests (see Figure 2.3), we can confirm that requests are increasing in the merit of the case. Theory, however, would predict a more steep increase, especially under the English rule. One potential explanation for the less pronounced increase is that participants under-exploit their bargaining position, and hence are relatively insensitive to merit-induced differences in the disagreement outcomes. This argument goes in line with the literature, who find that participants are relatively insensitive to changes in their disagreement values in the Nash-Demand game (Fischer et al., 2007; Anbarci and Feltovich, 2013), even if the bargaining position is earned through a preceding real-effort task (Anbarci and Feltovich, 2018).

Our key findings for the overall differences between the American and English rule are as follows. First, litigation expenditures are overall higher under the English rule compared to the American, including for low-merit cases. Second, we find no significant difference in settlement requests between the fee-shifting rules.

2.5.2 The Effect of Spite

In this subsection, we study how spiteful preferences affect litigation expenditures and settlement requests under both regimes and different merits.

Social Preferences and Behavior

First, we take a look at our measures of spiteful and prosocial attitudes. Even though this paper is about spiteful preferences, it is instructive to see whether the flip side of spite (i.e., prosociality) influences behavior in the opposite way. Thus, we use SVO as a robustness check throughout the entire results section. We find that the two spite measures are correlated positively and significantly ($r = 0.524$, $p<0.001$). Additionally, we see that our measure of prosocial behavior (SVO-Measure) is negatively correlated with our spite measure ($r = -0.132$, $p<0.001$) and with the Spite-Questionnaire ($r = -0.13$, $p<0.001$), providing plausibility for our measures of spiteful preferences.

Now we study the effect of spite on both litigation expenditures and settlement requests. As a first step, we correlate our measures of spiteful and prosocial preferences with litigation expenditures and settlement requests in linear regressions (see Table 2.2). Higher scores on the Spite-Task indicate stronger preferences for the destruction of wealth of the opponent, higher scores on the Spite-Questionnaire indicate more spitefulness, while increased social value orientation scores indicate more prosocial behavior. All independent variables are z-scored.

It can be seen that increasing spite scores (Spite-Task), as well as increasing spitefulness on the Spite-Questionnaire, are associated with higher legal expenditures and higher settlement requests. We also see that higher prosociality (SVO) is associated with lower settlement requests. An increase in the spite measures by one standard deviation influences legal expenditures and settlement requests more than a one standard deviation increase in prosociality.

That indicates that antisocial preferences play a more prominent role in describing behavior than prosocial preferences.

In the following, we rely on the behavioral Spite-Task as the main measure, but all the results can be replicated using the Spite Questionnaire. The results with SVO and the Spite-Questionnaire can be found in the appendix.

Table 2.2: Regression of litigation expenditures and settlement requests on social-preferences measures

	Dependent Variable:					
	Litigation / Settlement					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	5.26***	5.08***	5.26***	5.08***	5.26***	5.08***
	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.05)
Spite-Task	0.65***	0.68***				
	(0.05)	(0.05)				
SpiteQ			0.71***	0.83***		
			(0.05)	(0.04)		
SVO					−0.03	−0.11**
					(0.05)	(0.05)
Litigation	✓	×	✓	×	✓	×
Observations	1,635	1,635	1,635	1,635	1,635	1,635
R^2	0.09	0.12	0.11	0.18	0.0002	0.003
Adjusted R^2	0.09	0.12	0.11	0.18	−0.0004	0.003

Note: Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Standard errors are shown in parentheses. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Interaction of Spite with the Fee-Shifting Rule

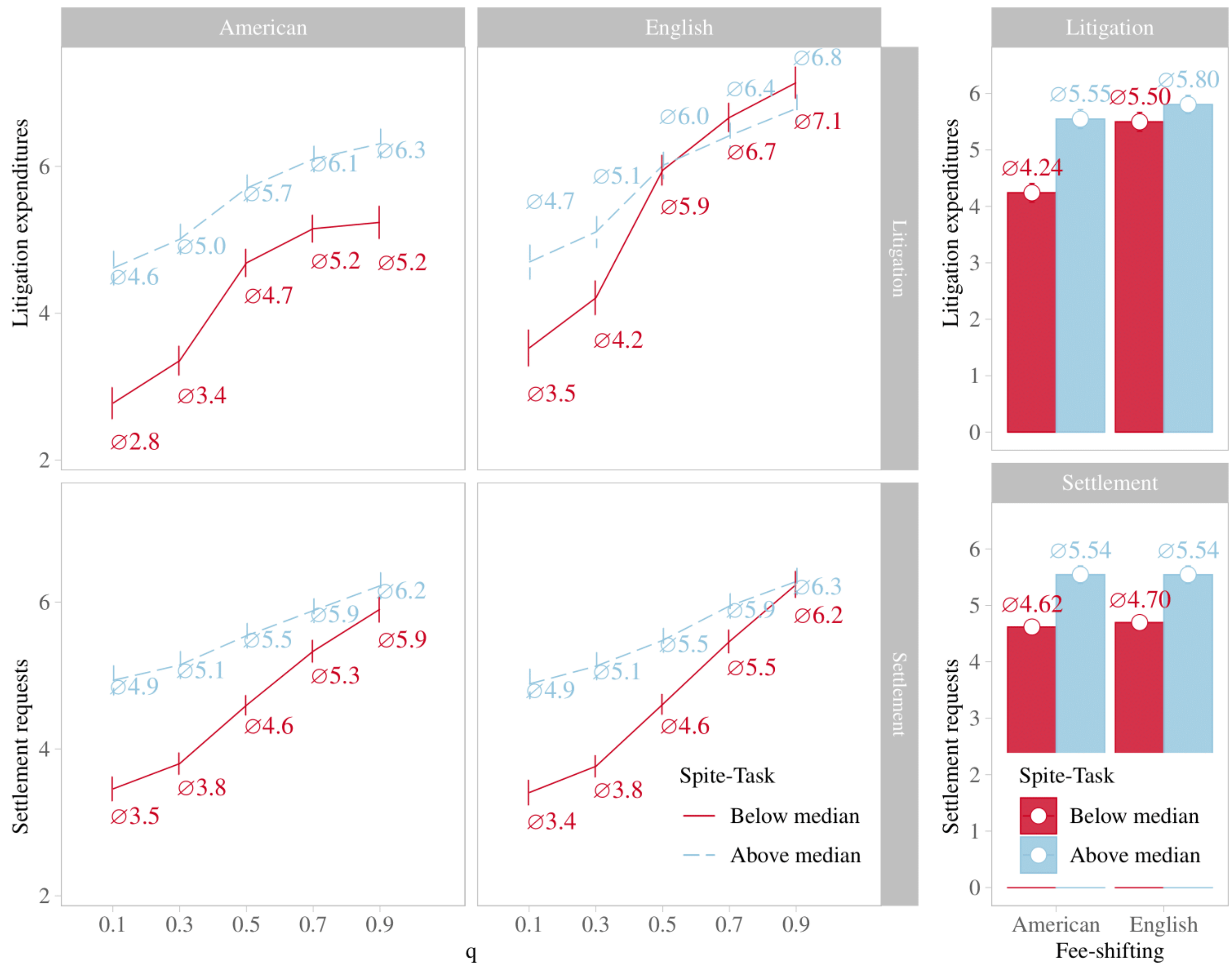
In the following, we classify subjects as spiteful if their spite score is higher than the median spite score and as non-spiteful otherwise, to obtain a deeper insight into the relationship between spite and litigation and settlement behavior depending on the fee-shifting rule and the merit of the case.[35] Figure 2.4 shows the litigation expenditures and the settlement requests for more and less spiteful subjects by the fee-shifting rule and the merit of the case (q).[36] To study these behavioral patterns formally, we use a mixed-effects regression for the aggregate legal expenditures and settlement requests by the median splits of the spiteful

[35] We caution that having a higher score than the median does not necessarily make a subject spiteful in absolute terms. However, we decided for this classification to have two balanced sets of subjects.

[36] In Appendix B.3.2 and B.3.3, we also show the settlement requests and the litigation expenditures for more and less spiteful subjects identified through the Spite-Questionnaire in Figure B.1 and for more and less prosocial subjects in Figure B.2.

preference measures and the fee-shifting rule (reported in Appendix B.3.2) and the merit of the case q (see Appendix B.3.3).[37] The following deductions from the figure are supported by the formal econometric analysis.

Figure 2.4: Litigation effort and settlement requests under the American and English fee-shifting rule as a function of q for more and less spiteful subjects



Note: The figures to the left depict the litigation expenditures and settlement requests by fee-shifting rule for more and less spiteful subjects as a function of q, while the figures to the right show the aggregates. The panels on the top show the litigation effort, while the panels on the bottom illustrate the settlement requests. Red solid lines depict the behavior of less spiteful subjects (i.e., subjects with below-median spite scores on the Spite-Task), while blue dashed lines indicate the response of more spiteful subjects. The error bars indicate the 95% confidence intervals.

Both under the English and American fee-shifting rule, we find that subjects with above-median spite scores (on both spite measures), on average, invest significantly more into litigation, as Hypotheses 1.2 and 1.3 suggest. Concerning settlement requests, we find that

[37]We also present the main results with continuous measures of social preferences in Appendix B.3.3. The results are essentially identical.

subjects with above-median spite scores, on average, request significantly higher settlement concessions under both rules.

When examining the interaction of spitefulness with the fee-shifting rule, we find that the increase in litigation expenditures for more spiteful subjects is significantly more pronounced under the American rule compared to the English rule. Thus, it appears that the American fee-shifting rule is more prone to distortions driven by spiteful preferences. As a consequence, litigation expenditures are relatively similar for more spiteful subjects under both rules. When examining this increase depending on the merit q, we see that the increase is relatively constant over all merit levels under the American rule, while under the English rule, it appears that more spiteful subjects exhibit substantially higher litigation expenditures for low-merit levels only. For high-merit levels, more and less spiteful subjects exhibit about the same litigation expenditures under the English rule. This goes tentatively in line with the functional predictions (see Figure 2.4), where, under the American rule, we predict a proportional increase for more spiteful agents for all merit levels, whereas, under the English rule, we predict an increase for a specific range of low-merit cases only.

Interestingly, we see no interaction effect between spiteful preferences and the fee-shifting rule for settlement requests. This indicates that spiteful preferences play roughly the same role under both fee-shifting rules for settlement requests. However, we do observe differential effects of spite depending on the merit of the case. We see that more spiteful subjects request substantially more than less spiteful subjects for low merits, and this difference decreases as the merit of the case increases. This pattern is found for both fee-shifting rules alike, and different to the theoretical functional predictions. For settlement requests, theory predicts no influence of spite under the English rule and an increase in litigation expenditures for high-merit cases and a decrease for low-merit cases.

This increase in settlement requests for both rules, with a decreasing difference, is consistent with the effects spite has in the bargaining stage. As pointed out in Section 2.5.1, literature tells us, that subjects systematically under-exploit their bargaining position (Fischer et al., 2007; Anbarci and Feltovich, 2013, 2018). This indicates that the effects of spite from the litigation stage, and thus from the disagreement outcomes, may receive less importance in the decision-making process. Consequently, the effects of spite in the bargaining stage may receive more importance, where any concession to the opponent creates additional disutility. Higher concessions create more spite-induced disutility, and thus, for low-merit cases, spite increases the disutility of the bargaining outcome more for the disadvantaged player compared to their opponent. Consequently, settlement requests increase. For high-merit cases, the concessions to the opponent are low, but one could anticipate the spiteful opponent wanting to concede less. Hence, in order to reach a settlement, the high-merit player has to decrease the requests. At the same time, the high-merit player wouldn't mind too much to litigate and consequently, still increases the request – but to a smaller extent than the low-merit player. This could explain, why we observe higher requests for low-merit cases and only smaller increases for high-merit cases due to spite. This reasoning is also underlined by the fact that the overall steepness of the settlement requests with respect to the merit q is less pronounced for players with above-median spite scores compared to players with below-median spite scores. In fact, this interpretation could also explain why it generally *seems* that players under-exploit their bargaining position: because they receive spite-related disutility directly from the bargaining outcomes, which attenuates the influence of their bargaining power.

The key findings of this section are: First, spiteful preferences are consistently associated with higher litigation expenditures and settlement requests under both rules. The increase in litigation expenditures is more pronounced under the American compared to the English rule, while for settlement requests there is no such differential effect depending on the fee-shifting rule. Consequently, the rules are relatively similar for more spiteful subjects, while for less spiteful subjects, the overall differences in litigation expenditures between the fee-shifting rules are even more pronounced.

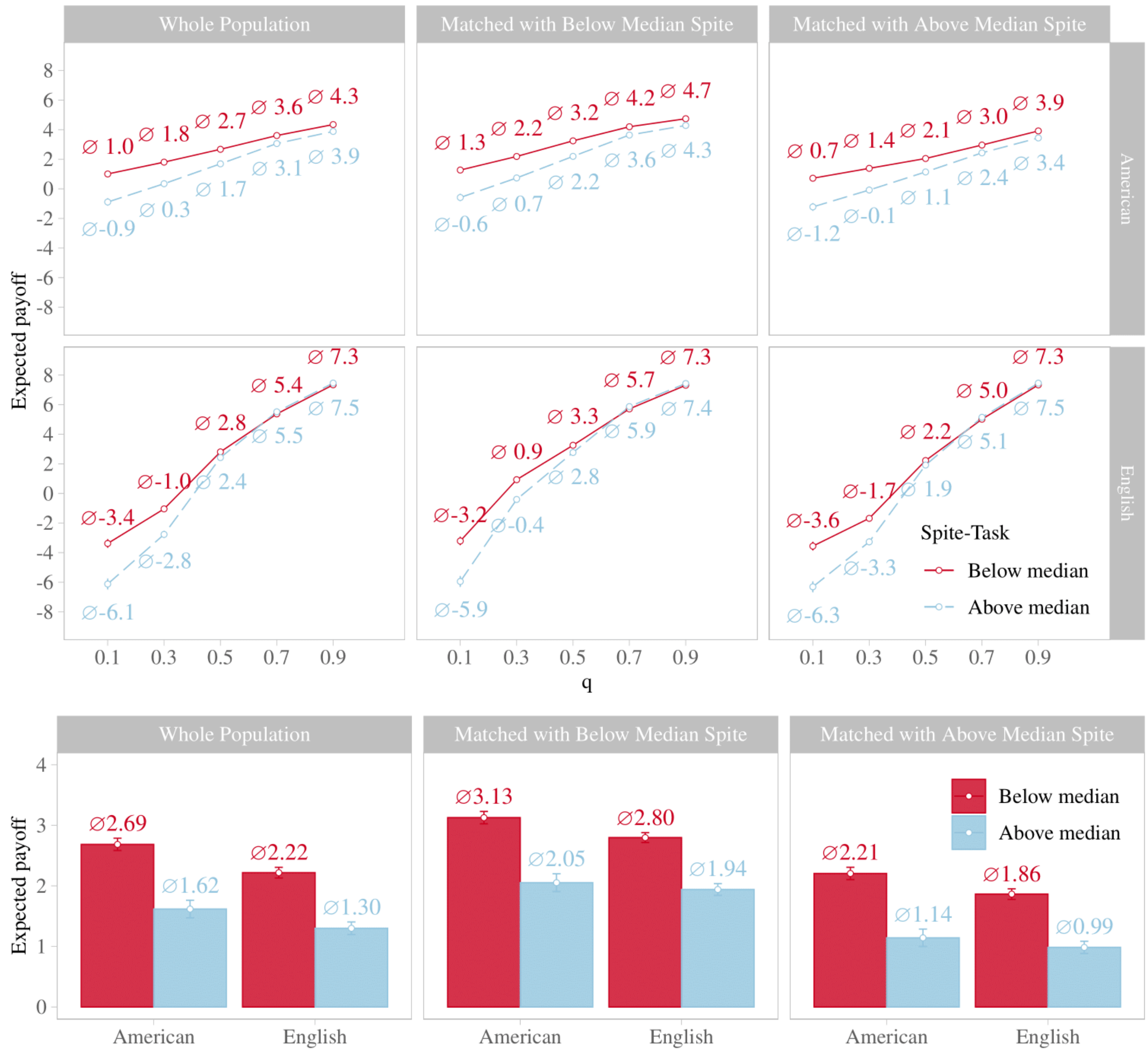
2.5.3 The Costs of Spite

In this section, we focus on the welfare implications of spitefulness depending on the fee-shifting rule and merit. Figure 2.5 depicts the empirical ex-ante expected payoffs for more and less spiteful subjects when matched with any kind of subject of the entire experiment's population under both fee-shifting rules.[38] It additionally shows the expected payoffs for being matched either with only more or only less spiteful subjects. We find that being spiteful comes at a considerable cost. Prior to bargaining and litigation, less spiteful subjects are expected to have an average payoff of 2.69 tokens compared to an average payoff of 1.62 tokens for more spiteful subjects under the American rule ($t(1633)= 12.0$, $p < 0.001$), and 2.22 tokens compared to 1.30 tokens under the English fee-shifting rule ($t(1633)= 13.3$, $p < 0.001$).

This decrease in expected payoffs for more spiteful subjects is independent of who they are matched with. The bar charts of Figure 2.5 show that the empirical ex-ante expected payoffs of more spiteful subjects are always lower compared to less spiteful subjects, both when they are matched only with above median spite subjects or only with below median

[38]The expected payoff is the payoff a subject with a given merit q is expected to obtain prior to bargaining and litigation. We focus on the expected payoff and not on expected utility to measure the monetary costs that spitefulness bears to society.

Figure 2.5: Expected payoff by fee-shifting rule as a function of q



Note: The panels on the top depict the expected payoff by fee-shifting rule as a function of q, while the panels on the bottom show the aggregates. The panels in the first column show the expected payoffs of more (in blue) or less (in red) spiteful subjects when being matched with any of the other subjects, while the second and third columns illustrate the expected payoffs when being matched either only with less or more spiteful subjects. The error bars indicate the 95% confidence intervals.

spite subjects. All shown differences are highly significant at the 0.1% level using t-tests. Hence, in monetary terms, it does not pay off to be spiteful.[39] This has two reasons, which persist independent of who spiteful players are matched with. First, there is only a slight increase in winning chances for more spiteful players (see Figure B.6 in Appendix B.6), which does not fully compensate for the increase in litigation expenditures. Second, settlement

[39]We acknowledge that more spiteful players receive additional non-monetary utility from winning, which could offset their monetary losses.

probabilities decrease for more spiteful subjects (see Figure B.7 in Appendix B.6), making litigation and the associated costs more likely.

Next, we compare the difference in this decrease between the American and English fee-shifting rule. Being matched with the entire population, the decrease in the expected payoffs for being spiteful is more pronounced under the American compared to the English fee-shifting rule (-1.07 tokens vs. -0.92 tokens). This difference in the strength of the penalty is driven by differential effects of the fee-shifting rule depending on the merit of the case. Under the American rule, being more spiteful comes at a cost for all merit cases, whereas under the English rule, more spiteful subjects obtain lower expected payoffs for low-merit cases only. Again, these dynamics are independent of being matched either to below or above median spite subjects, or with the entire experiment's population. The regression analysis in Appendix B.3.4, Table B.6 reveals that this difference in the decrease is always significantly (at least at the 10% level) more pronounced under the American rule. Consequently, the American rule punishes players more for their spitefulness than the English rule.

But not only more spiteful subjects obtain lower expected payoffs. The existence of spiteful subjects is detrimental to less spiteful subjects, too. The average expected payoff of below-median-spite subjects is 3.13 tokens and 2.80 tokens when being matched with other less spiteful subjects under the English and American rule, respectively, whereas it significantly decreases to 2.21 tokens and 1.86 tokens when matched with above median spite subjects ($t(3268)= 19.3$, $p < 0.001$ and $t(3268)= 13.9$, $p < 0.001$). This difference in the expected payoff is not significantly different between the two fee-shifting rules (see Appendix B.3.4, Regression Table B.7). Hence, none of the two rules can protect less spiteful players better from the harm that more spiteful players inflict on them.

We conclude that spitefulness does not only reduce payoffs for subjects displaying spiteful behavior but also for less spiteful subjects, who are matched with more spiteful participants. The American rule punishes being spiteful more than the English rule by having a more pronounced negative impact on their expected payoffs. However, the inflicted harm upon less spiteful subjects does not significantly differ between the rules.

2.6 Discussion and Conclusion

In this paper, we study how spiteful preferences change behavior in litigation settings under the American compared to the English fee-shifting rules. We show theoretically that spiteful preferences lead to higher litigation expenditures under both rules. For settlement requests, spite matters only under the American fee-shifting rule, where it increases requests for low-merit cases and decreases requests for high-merit cases.

Using an online experiment, we provide empirical evidence for some of these predictions. In the experiment, subjects had to make litigation and settlement decisions under the American and English fee-shifting rules. We elicited spiteful preferences via two measurements, namely through 1) a behavioral incentivized distribution-decision task and 2) a non-behavioral questionnaire.

We find that litigation expenditures are overall higher under the English than under the American fee-shifting rule. This goes in line with the experimental results of Dechenaux and Mancini (2008), Gabuthy et al. (2021), and Massenot et al. (2021). We extend their finding by exogenously varying the merit of a case. In the theoretical literature, is often argued that the English rule has the advantage of deterring low-merit frivolous lawsuits (Spier, 2007). We do not find any evidence for this claim. Instead, we even observe higher litigation expenditures under the English rule for all merit levels, including for low-merit ones. This result goes in line with Gabuthy et al. (2021), who find that the English rule can increase low-merit cases. Unlike Dechenaux and Mancini (2008) and Gabuthy et al. (2021), but similar to Massenot et al. (2021), we find no difference in settlement requests and rates between the two fee-shifting rules. As a consequence, the English fee-shifting rule does not seem to deter low-merit cases but rather increases litigation expenditures for such frivolous lawsuits. Not only for frivolous lawsuits but also overall, the English fee-shifting rule leads to the use of more judicial resources compared to the American rule, which may lead to welfare losses as those resources can not be used otherwise. This conclusion, however, depends on whether litigants exhibit spiteful preferences, gaining additional utility from inflicting harm on their opponents.

The main insight of this paper is that litigation and settlement expenditures are higher under both fee-shifting rules if subjects exhibit such spiteful preferences – with a more pronounced increase under the American rule compared to the English rule. Consequently, litigation expenditures – and thus the use of judicial resources – are only slightly higher under the English rule. For less spiteful subjects, however, the earlier conclusion prevails, as the English rule leads to substantially higher litigation expenditures for all merit levels. The degree, to which the English rule leads to the use of more judicial resources compared to the American thus depends on the fraction of spiteful litigants.

The more pronounced increase in litigation expenditures under the American rule goes in contrast to the theoretical results of Chen and Rodrigues-Neto (2023), who show that negative relational emotions amplify the cost-shifting effect, yet in a slightly different theoretical model. We find that this more pronounced increase is driven by a constant increase due to spite under the American rule for all merits, whereas under the English rule, there is an increase for low-merit cases only. This goes in line with our theoretical model, which predicts only an increase for a specific range of low-merit levels and no difference for more spiteful agents for high-merit levels.

Concerning pre-trial bargaining, we find that more spiteful subjects demand higher settlement requests under both fee-shifting rules. This increase does not significantly differ between the two rules, and hence spite seems to not have a stronger influence between the two fee-shifting rules, contrary to our theoretical prediction. Additionally, spite also does not have an opposite effect depending on the merit of the case under the American rule. Instead, spite increases settlement requests for all merit levels under both rules, yet less for high-merit cases. This suggests that subjects do not (fully) exploit their bargaining power, which is shaped by the expected litigation outcomes. Rather, it appears that spiteful subjects care more about the settlement outcome itself – where any concession to the opponent creates additional spite-driven disutility – and demand overall higher requests. This tendency is particularly notable for low-merit cases, where they initially concede a lot to their opponent and thus this tendency provides a plausible explanation for our observations. Previous experimental work also finds that subjects are relatively insensitive to changes in the disagreement values (Fischer et al., 2007; Anbarci and Feltovich, 2013, 2018). We propose an explanation of why this might be the case: because subjects receive spite-induced disutility from any concession to the opponent, which directly attenuates the bargaining power derived from the disagreement values.

Finally, overall higher demands lead to more settlement failures. Consequently, lawsuits, including frivolous ones, are more likely to go to court, where players motivated by spiteful preferences also invest more in litigation.

By showing an overall increase in litigation expenditures and settlement requests for more spiteful subjects under both rules, we complement the experimental literature (Kimbrough and Reiss, 2012; Cooper and Fang, 2008; Bartling et al., 2017; Andreoni et al., 2007; Kirchkamp and Mill, 2021a), which shows that spiteful preferences lead to more competitive behavior and to the results of Eisenkopf et al. (2019), who do not find any impact of emotions on litigation expenditures. That could indicate that in case of litigation, antisocial preferences matter more than 'hot' negative emotions.

We also study the harm that is caused because of spitefulness. For that, we compare the expected payoffs of more spiteful compared to less spiteful subjects. We find that it does not pay off to be spiteful, as the expected payoff is lower for more spiteful subjects independent of being matched with either only less or more spiteful subjects. This decrease in the expected payoff is more pronounced under the American compared to the English rule. The English rule, therefore, protects spiteful players better from decreasing their own monetary payoffs. The harm inflicted upon others, however, does not differ between the two rules. As a consequence, the fee-shifting rule can not be used to mitigate the harm that players suffer from being matched with a spiteful opponent.

We conclude that spiteful preferences are shown to be bad news – not only are litigation expenditures and settlement requests higher for more spiteful preferences, but also the expected payoffs are lower both for spiteful players and their matched partners. Neither of the fee-shifting rules can protect the harm that spiteful players inflict on others. The English rule, however, can decrease the penalty that spiteful players receive – yet at the cost of yielding overall higher litigation expenditures, especially for less spiteful players, including an increase in expenditures for low-merit cases. The choice of the fee-shifting rule hence depends on the distribution of spiteful players and the merits of the case.

There are some limitations of the study, which the reader should take into account. First, we choose a specific simplified version of spite in our theoretical model. While we rely mostly on the existing literature to formulate spiteful preferences, there are many possible alternative ways of modeling spite. Future research might want to tackle this limitation by focusing on broader models of spiteful preferences. Second, we rely on one-shot interactions in our experiment setting. This approach does not leave room for learning. In many experimental contest settings, learning plays a crucial role in behavior changes over time (see e.g., March and Sahm (2017)). At the same time, experiments with repeated interactions might fail to attribute changes in behavior to preferences. We cannot answer how participants would learn and how this learning would interact with spite and the fee-shifting rule. We can, however, show that participants with higher spiteful preferences differ already substantially from participants with less spiteful preferences in a one-shot setting. Thus, it would seem plausible that our results would even exacerbate over time. Third, we elicit litigation expenditures for all subjects independent of whether they settle or actually have to litigate. Hence, we shut down selection effects for the litigation stage. In reality, there exists a selection effect, in the sense that only subjects that fail (or do not want) to settle, litigate. We purposefully excluded this selection effect to keep our results clean. A selection effect most likely would even magnify our results (as the more spiteful litigants would be less likely to settle) and thus, it seems plausible that the effect of spite is even stronger in real-world settings of litigation. Finally, even though the comparison between the American and English rules is causal (due to an exogenous treatment manipulation), we do not exogenously manipulate spite. We, instead, rely on correlational evidence on the influence of spite on behavior. The main reason for not exogenously varying spiteful preferences is that we are not aware of any manipulation, which cleanly targets only spiteful preferences while keeping other preferences and beliefs constant. However, we tackle this issue throughout the paper (see Appendix B.5 for a discussion of causality). To prevent the results from being driven by measurement error, we elicit spiteful preferences via two different methods. Throughout the paper, we consistently show that all our results prevail using either measure of spiteful preferences (see also Appendix B.3.2 and B.3.3). Further, we tackle a potential omitted-variable bias problem by running robustness checks with risk preferences and other controls that may be correlated both with the spite measurement and litigation expenditures and settlement requests (see Appendix B.4). The results remain robust for these additional model specifications. Even though all these results make us rather confident that spiteful preferences indeed change litigation and settlement behavior, we cannot exclude the possibility of reversed causality or omitted-variable bias. Thus, future research might want to find ways of cleanly manipulating only spiteful preferences to be able to provide causal evidence to our research question.

Notwithstanding the limitations of our study, future research could explore alternative mechanisms to mitigate excessive litigation driven by spiteful preferences. For instance, research could investigate whether the negative effects of spiteful preferences on litigation and settlement can be alleviated through lawyers and contingency fees or potential cool-off periods.

All in all, we consistently find that spiteful preferences are associated with higher litigation expenditure and settlement requests, which result in welfare losses. We find that the English fee-shifting rule is more robust towards spiteful preferences and thus protects spiteful litigants more from decreasing their expected payoffs compared to the American rule. However, the harm inflicted upon others is not different between the two rules. Additionally, the English rule leads to higher litigation expenditures overall, including for frivolous low-merit lawsuits. Therefore, the American rule seems to be welfare-improving compared to the English rule, as we find that it leads to the use of fewer judicial resources including for frivolous low-merit lawsuits – especially so when players are not much motivated by spiteful preferences.

Acknowledgments for Chapter 2

We thank Kathryn Spier, Oliver Kirchkamp, Henrik Orzen, Jonathan Klick, Sven Höppner, Serhiy Kandul, Olexandr Nikolaychuk, Leonard Hoeft, Benedikt Werner, and Christoph Engel for helpful comments. We appreciate comments from participants of the IMPRS Thesis Workshop, the Econometric Society European Meeting in Manchester, the Ghent Law & Econ Seminar, the Jena Econ-Seminar, the YEM in Brno, the 8th Annual Conference on "Contests: Theory and Evidence" in Reading, and the HeiKaMaxY in Mannheim. We gratefully acknowledge funding from the Max Planck Society through the IMPRS-Uncertainty. Support by the German Research Foundation (DFG) through CRC TR 224 (Project A01) is gratefully acknowledged. This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences.

Appendix B

Appendix to Chapter 2

B.1 Propositions

B.1.1 Litigation Model Proposition

Proposition 1. *The symmetric litigation expenditures under the American fee-shifting rule for spiteful agents are given by:*

$$e^{*(\mathrm{Am})}(W, q, \alpha) = (1-q) \cdot q \cdot W \cdot (\alpha + 1)$$

The proof of Proposition 1 is shown in Appendix B.2.1. The equilibrium litigation expenditures under the English fee-shifting rule are given below.

Proposition 2. *The litigation expenditures under the English fee-shifting rule for spiteful agents are given by:*

$$e_i^* = \begin{cases} W & \text{if } q \leq \overline{q}(\alpha) \\ 0 & \text{else} \end{cases} \qquad \text{with } \overline{q}(\alpha) = \frac{1}{3}\frac{3\alpha+2}{(\alpha+1)},$$

$$e_j^* = \begin{cases} 0 & \text{if } q \leq (1-\overline{q}(\alpha)) \\ W & \text{else} \end{cases} \qquad \text{and } (1-\overline{q}(\alpha)) = \frac{1}{3}\frac{1}{(\alpha+1)}$$

The proof of Proposition 2 is shown in Appendix B.2.2. Figure 2.1 shows the equilibrium behavior for player j for different levels of α.Litigation expenditures under the English fee-shifting rule are characterized by the bang-bang property. For low merits, it is optimal to incur no expenditures, and after a certain threshold, it is optimal to incur full expenditures. For more spiteful agents, this threshold is shifted towards lower merit levels. More spiteful agents, thus, incur full expenditures at lower merit levels than less spiteful agents.

A rough interpretation is the following: Under the English fee-shifting rule, the loser has to carry the costs from both parties. Therefore, the disutility from losing and the utility of winning is augmented compared to the American rule. Hence, for high-merit cases, it is optimal to incur full expenditures because they very likely do not have to be paid by the winning party. This decreases the winning probabilities for low-merit cases further and hence, it is optimal to reduce own expenditures to the minimum as they have to be carried almost certainly by oneself. At the threshold, the augmented incentives to win outweigh the costs of potentially paying their own expenditures.[1] After the threshold, it is optimal to incur full expenditures.[2]

More spiteful agents have even more augmented incentives to win since they receive additional disutility from losing (since the opponent has a positive payoff) and additional utility from winning (since the opponent has to carry all the costs). Therefore, the threshold to switch from no expenditures to full expenditures moves to lower merit levels. Note that for small enough merit levels $q < (1 - \overline{q}(\alpha))$, the expected utility of a player is negative since losing means carrying both costs (either W or 2W) and winning means receiving W. Therefore, having to litigate under the English fee-shifting rule is bad news if the own merit is not high enough.

B.1.2 Settlement Model Proposition

Spite shapes this equilibrium outcome simultaneously through the settlement stage and the litigation stage. There are two countervailing forces. First, in the litigation stage, spite influences the disagreement values (i.e., the expected payoffs). If one of the players wins the litigation, spite increases her utility of litigation. If this player loses, spite lowers her utility. In the equilibrium outcome, for low-merit levels ($q < 0.5$), player j has a winning probability of less than 0.5. Therefore, spite decreases player j's expected utility further compared to player i's. Subsequently, a spiteful player j is less eager to litigate than a spiteful player i in the low-merit case, and hence, player j's bargaining power decreases. More spiteful agents then request less in the settlement stage for low-merit casesthan less spiteful agents. Due to symmetry, requests are higher for high-merit cases ($q > 0.5$).

Second, in the settlement stage, spite interacts with the opponent's demands and creates a countervailing force. For low-merit levels ($q < 0.5$), player j's expected utility of the

[1]Due to the convex form of the utility function, which has the minimum utility level in between 0 and W expenditures, only a switch to full expenditures maximizes the expected utility. A partial increase in the expenditures would not increase the probability of winning enough to counterbalance the increased costs.

[2]This bang-bang property can be illustrated best with an example: Suppose that $q = \frac{1}{3}$ and $W = 10$. At this merit level, a non-spiteful player ($\alpha = 0$) is indifferent between spending 0 and 10 because the expected payoff is the same. For exerting 0, the player loses with certainty, and the utility is $u_j = -10$ because of the opponent's expenditures of 10. Spending 10, while the opponent also spends 10, leads to a winning probability of $\frac{1}{3}$ for the prize of 10. Hence, the expected utility is $E[u_j] = \frac{1}{3} * 10 - \frac{2}{3} * (-20) = -10$. If the merit level is slightly below $\frac{1}{3}$, spending 0 maximizes the expected utility. If it is slightly above $\frac{2}{3}$, spending 10 maximizes the expected utility.

litigation stage is smaller than player i's, and hence she has a smaller bargaining power than player i. Subsequently, player i's demands are higher than player j's. Higher demands of the opponent are associated with a higher disutility due to spite. Hence, a more spiteful player j has a higher disutility due to spite for low-merit cases than a less spiteful player. Therefore, player j is less eager to settle at these conditions and her bargaining power increases compared to player i's. Subsequently, requests in the settlement stage are higher for more spiteful agents for low-merit levels ($q < 0.5$). Correspondingly, requests are lower for high-merit levels ($q > 0.5$) because of symmetry. Whether the first or second effect prevails depends on the payoff structure and environment that is determined either by the American or English fee-shifting rule.

Proposition 3. *Under the American fee-shifting rule, the requests of players i and j are characterized by the following functions:*

$$s_i^* = W - s_j^* = \begin{cases} W & \text{if } q \leq \frac{1}{2}\frac{\alpha}{\alpha+1} \\ W(\alpha(\frac{1}{2} - q) + (1-q)) & \text{if } \frac{1}{2}\frac{\alpha}{\alpha+1} < q < \frac{1}{2}\frac{\alpha+2}{\alpha+1} \\ 0 & \text{if } q \geq \frac{1}{2}\frac{\alpha+2}{\alpha+1} \end{cases}$$

The proof can be found in Appendix B.2.3. Figure 2.2 shows the equilibrium settlement requests for player j under the American rule. For non-spiteful agents, there is a linear and constant increase in the requests with increasing merit of the case ($\lim_{\alpha\to 0} s_j^* = Wq$) since the outside value and hence the bargaining power increases. For lower merit levels ($q < 0.5$), more spiteful agents request less, whereas, for higher merit levels ($q > 0.5$), more spiteful agents request more. This is because more spiteful agents want to prevent litigation if their merit is low and wouldn't mind litigating when their merit is high since the outside values are augmented. Therefore, under the American rule, spite in the disagreement values outweighs the effect of spite in the settlement requests.

Proposition 4. *Under the English fee-shifting rule, the requests of player i and j are characterized by the following functions:*

$$s_i^* = W - s_j^* = \begin{cases} W & \text{if } q \leq \frac{1}{3} \\ (2-3q)W & \text{if } \frac{1}{3} < q < \frac{2}{3} \\ 0 & \text{if } q \geq \frac{2}{3} \end{cases}$$

The proof can be found in Appendix B.2.4. Figure 2.2 depicts the equilibrium settlement requests under the English fee-shifting rule. Requests start at 0 for low-merit cases, then increase after the merit is at $q = \frac{1}{3}$ until $q = \frac{2}{3}$, where they stay at the maximum request W. Notice that requests are the same for all spite levels. This is because the opposing effects of spite in the disagreement values and spite in the settlement requests cancel each other out.

B.2 Proofs

B.2.1 Proof of Proposition 1

As a reminder, the utility of player i looks as follows (equation 2.1):

$$u_i(e_i, e_j, q, \alpha) = \underbrace{p_i(e_i, e_j, q)\cdot}_{\text{Probability of winning}} \underbrace{\left(\underbrace{W - 1_{American} e_i}_{\text{Payoff}} \underbrace{- \alpha \cdot (-e_j - 1_{English} e_i)}_{\text{Disutility due to spite}}\right)}_{\text{Utility in case of winning the case}}$$

$$+ \underbrace{(1 - p_i(e_i, e_j, q))\cdot}_{\text{Probability of losing}} \underbrace{\left(\underbrace{-e_i - 1_{English} e_j}_{\text{Payoff}} \underbrace{- \alpha \cdot (W - 1_{American} e_j)}_{\text{Disutility due to spite}}\right)}_{\text{Utility in case of losing the case}}$$

Proof of Proposition 1. Differentiating the above equation with respect to e_i gives:

$$\frac{\partial u_i(e_i, e_j, q, \alpha)}{\partial e_i} = \frac{\left(\left((\alpha+1)1_{English} - 1\right) e_i{}^2 - 2e_j \left((\alpha+1)\, 1_{English} - 1\right) e_i - e_j \left(\left(1 + (\alpha+1)\, 1_{English}\right) e_j + W\,(\alpha+1)\right)\right) q^2}{\left((e_i - e_j)q - e_i\right)^2}$$
$$+ \frac{\left(2\left((\alpha+1)\, 1_{English} - 1\right) e_i (e_j - e_i) + e_j\,(\alpha+1)\left(1_{English} e_j + W\right)\right) q + \left((\alpha+1)\, 1_{English} - 1\right) e_i{}^2}{\left((e_i - e_j)q - e_i\right)^2}$$

The second derivative is given by:

$$\frac{\partial \frac{\partial u_i(e_i, e_j, q, \alpha)}{\partial e_i}}{\partial e_i} = 2\frac{(\alpha+1) e_j (q-1) q \left((2 1_{English} e_j + W)q - 1_{English} e_j - W\right)}{\left((e_i - e_j)q - e_i\right)^3} \tag{B.1}$$

Rearranging yields the best response for agent i given a merit q, a good W, spite α and the litigation expenditures of j:

$$e_i^*(e_j, q, \alpha, 1_{English}, W) = \frac{1}{1-q}\left(-qe_j \pm \sqrt{\frac{qe_j\ (\alpha+1)\left((-1+q)\, W + (2\,q - 1)\, e_j\, 1_{English}\right)}{-1 + (\alpha+1)\, 1_{English}}}\right)$$

From the best response function, we can derive the equilibrium behavior. As we know that the best response of j is given by $e_j^*(e_i, q, \alpha, W) = e_i^*(e_i, 1-q, \alpha, W)$, we insert the best response of j into the best response of i. We obtain:

$$e_i^{*(\mathrm{Am})}(e_j, q, \alpha, W) = (1-q) \cdot q \cdot W \cdot (\alpha + 1)$$

The second derivative (Equation B.1) yields:

$$\frac{\partial \frac{\partial u_i((1-q)\cdot q\cdot W\cdot(\alpha+1),(1-q)\cdot q\cdot W\cdot(\alpha+1),q,\alpha)}{\partial e_i}}{d\,e_i} = \frac{-2}{W(\alpha+1)q}$$

which is negative and hence, the solution is maximizing the utility of i. □

B.2.2 Proof of Proposition 2

Proof of Proposition 2. The utility function under the English fee-shifting rule can be rewritten as follows:

$$U_i = \frac{(1-q)e_i}{(1-q)e_i + qe_j}[(1+\alpha)e_i + (1+\alpha)e_j + (1+\alpha)W] - e_i - e_j - \alpha W$$

Unlike under the American fee-shifting rule, there are self-generated spillovers under the English regime because own expenses increase the value of the winning prize by generating spite-driven utilities, which are determined by the spite parameter α. In an unconstrained optimization and best response equilibrium, both infinite expenses and negative expenses are employed. Therefore, we employ a constrained optimization. With the constraints, we restrict the possible resources spent and prevent the agents from spending infinite and negative resources. In addition to guaranteeing mathematical solvability, constraining effort levels also reflect reality since agents do not have infinite resources and cannot exert negative efforts. We set $\overline{e} = W$, i.e., agents are spending at most the value of the prize. The constrained optimization problem looks as follows:

$$\begin{aligned} &\max_{e_i} U_i \\ \text{s.t. } &e_i \leq W \\ &e_i \geq 0 \end{aligned}$$

The point (e_i^*, μ^*) is called a Karush-Kuhn-Tucker (KKT) point if the following equations hold:

$$\begin{aligned} \frac{\partial U_i(e_i^*)}{\partial e_i} - \mu_1(\frac{\partial g_1(e_i^*)}{\partial e_i^*}) - \mu_2(\frac{\partial g_2(e_i^*)}{\partial e_i^*}) &= 0 \\ g_1(e_i^*) = -e_i^* &\leq 0 \\ g_2(e_i^*) = e_i^* - W &\leq 0 \\ \mu_1 &\geq 0 \end{aligned}$$

$$\mu_2 \geq 0$$
$$\mu_1 g_1(e_i^*) = 0$$
$$\mu_2 g_2(e_i^*) = 0$$

We obtain the following points that may satisfy the KKT conditions for specific values of the parameters.

$$(e_i^* = 0, \mu_1 = \frac{e_j q - e_j(1-q)(1+\alpha) - W(1-q)(1+\alpha)}{qe_j}, \mu_2 = 0) \tag{B.2}$$

$$(e_i^* = W, \mu_1 = 0, \mu_2 = \frac{e_j^2 q(1-q)(1+\alpha) - e_j^2 q^2 + e_j qW(3\alpha+1)(1-q) + W^2\alpha(1-q)^2}{((1-q)10 + qe_j)^2}) \tag{B.3}$$

$$(e_i^* = \frac{1}{1-q}(-e_j q + \sqrt{\frac{qe_j(\alpha+1)[(-1+q)W + (-1+2q)e_j]}{\alpha}}), \mu_1 = \mu_2 = 0) \tag{B.4}$$

$$(e_i^* = \frac{1}{1-q}(-e_j q - \sqrt{\frac{qe_j(\alpha+1)[(-1+q)W + (-1+2q)e_j]}{\alpha}}), \mu_1 = \mu_2 = 0) \tag{B.5}$$

The optimization problem $e_j(q)$ equals the optimization problem for $e_i(1-q)$ and all the following conditions also apply to the optimization problem for player j. The sets $e_i \in [0, W]$ described by $g_1(e_i^*)$ and $g_2(e_i^*)$ are convex. Furthermore, the functions $g_1(e_i^*)$ and $g_2(e_i^*)$ are linear and affine. Therefore, they satisfy the linearity constraint qualification and thus all regularity conditions. The parameters are defined as before as $a \in (0,1), q \in [0,1], W \in (0, \infty)$ and $e_j \in [0, W]$. There is a region for a specific range of q, α, and W, where Point (B.2) and Point (B.3) are feasible, since both $\mu_1, \mu_2 \geq 0$ in Point (B.2) and $\mu_1, \mu_2 \geq 0$ in Point (B.3). Point (B.4) and Point (B.5) always satisfy $\mu_1, \mu_2 \geq 0$, yet there are some conditions on the parameters for the square root to be non-negative and e_i to be non-negative.

First, note that for $q = 0$, only points (B.4) and (B.5) are feasible and yield the optimal solution of $e_i = 0$. In the following, we analyze the optimal solution for $q \in (0, 1]$.

For $q \leq 0.5$, the maximum is at $e_i^* = W$. Point (B.2) is not feasible, since $\mu_1 < 0$ for $q \in (0, 0.5]$ and hence the necessary condition for an extreme point is not met. For $q \leq 0.5$ point (B.3) is feasible, since $\mu_1, \mu_2 \geq 0$ for $q \in (0, 0.5]$ and all the other parameters in their domain. Additionally, $\frac{\partial U_i}{\partial e_i} > 0$, $\forall q \in (0, 0.5]$ and all the other parameters in their domain and $\forall e_i \in [0, W]$, giving a sufficient condition for $e_i = W$ to be a maximizer for $q \leq 0.5$. This can be seen by the following;

$$\frac{\partial U_i}{\partial e_i} = \frac{(1-q)((1+\alpha)e_i + (1+\alpha)e_j + (1+\alpha)W)}{(1-q)e_i + qe_j} -$$

$$\frac{(1-q)^2 e_i((1+\alpha)e_i + (1+\alpha)e_j + (1+\alpha)W)}{((1-q)e_i + qe_j)^2} + \frac{(1-q)e_i(1+\alpha)}{((1-q)e_i + qe_j)} - 1 > 0$$

After some algebra it becomes:

$$qe_j(1-q)[2e_i\alpha + (1+\alpha)e_j + (1+\alpha)W] + (1-q)^2 e_i^2 \alpha - q^2 e_j^2 > 0$$

Note that the first and second terms are always positive and the third term is always negative. Further, the third term is always smaller than the first one for $q \in (0, 0.5]$. Therefore, for $q \leq 0.5$, more effort is always better, and hence, the maximum effort possible, $e_i = W$, is the optimal solution.

Since $e_j^*(q) = e_i^*(1-q)$, the best response from player j is always W independent of player i's action for $q \geq 0.5$. With that knowledge, we now describe the best responses of player i for $q > 0.5$, knowing that player j always exerts effort of W.

With that knowledge, Point (B.5) is never feasible, since for $q < \frac{2}{3}$, the square root is negative, and further the point always yields $e_i < 0$ for $q > 0.5$ and the respective $e_j^* = W$. Similarly, Point (B.4) is also not feasible for $q < \frac{2}{3}$, however, for $q \geq \frac{2}{3}$, there exist combinations of the parameters that yield a feasible solution.

The first region is for $q \in (0, \overline{q}_{\mu_1}(\alpha))$, where Point (B.3) is feasible and where Point (B.2) is just not feasible yet. $\overline{q}_{\mu_1}(\alpha)$ follows from setting $\mu_1 = 0$ from Point (B.2):

$$\overline{q}_{\mu_1}(\alpha) = \frac{2(\alpha+1)}{2\alpha+3} \in (\frac{2}{3}, \frac{4}{5})_{|\alpha \in (0,1)}$$

In this region both Point (B.3) and Point (B.4) are feasible. Since $\frac{\partial U_i}{\partial e_i} > 0, \forall q \in (0, \overline{q}_{\mu_1}(\alpha)], \forall e_i \in [0, W]$ and $e_j^* = W_{|q \geq 0.5}$, $e_i^* = W$ is the local maximizer for the whole first region.

The second region is for values of $q \in [\overline{q}_{\mu_1}(\alpha), \overline{q}_{\mu_2}(\alpha)]$, where the points (B.2),(B.3) and (B.4) are feasible. $\overline{q}_{\mu_2}(\alpha)$ is the threshold where the point (B.3) is just still feasible, so where $\mu_2 = 0$ from the point (5):

$$\overline{q}_{\mu_2}(\alpha) = \frac{1}{3} \frac{(\alpha+1) + \sqrt{4\alpha^2 + 5\alpha + 1}}{(\alpha+1)} \in (\frac{2}{3}, 0.86)_{|a \in (0,1)}$$

At this region, the utility function is convex in e_i. We compare the utility of the feasible points to get the local maximum of the region. Note, that by the convexity of the function in this region, one of the points is the minimum. By comparing the values, we find that Point (B.4) is the minimum in this region. Further, $U_i(e_i = W, e_j = W, \alpha, q = \overline{q}_{\mu_1}(\alpha), W) > U_i(e_i = 0, e_j = W, \alpha, q = \overline{q}_{\mu_1}(\alpha), W)$ and $U_i(e_i = W, e_j = W, \alpha, q = \overline{q}_{\mu_2}(\alpha), W) < U_i(e_i = 0, e_j = W, \alpha, q = \overline{q}_{\mu_2}(\alpha), W)$, indicating that throughout the region the best response changes from $e_i^* = W$ to $e_i^* = 0$. Because of the convexity of the utility function in this region, we find the bang-bang property, meaning that there exists a threshold $\overline{q}(\alpha)$, where the best response jumps from W to 0. We compute this threshold by equalizing the utilities from the two points: $U_i(e_i = W, e_j = W, \alpha, q, W) = U_i(e_i = 0, e_j = W, \alpha, q, W)$ and receive:

$$\overline{q}(\alpha) = \frac{1}{3}\frac{3\alpha+2}{\alpha+1} \in (\frac{2}{3}, \frac{5}{6})_{|a\in(0,1)}$$

We therefore showed that $e_i^* = W$ for $q \in (0, \overline{q}(\alpha)]$ and $e_i^* = 0$ for $q \in [\overline{q}(\alpha), \overline{q}_{\mu_2}(\alpha)]$ are local maximizers.

For the third region, $q \in (\overline{q}_{\mu_2}(\alpha), 1]$, it remains to show that $e_i^* = 0$ is a local maximizer. Note that for this region, Point (B.3) and Point (B.4) are feasible. Since, $\frac{\partial U_i}{\partial e_i} < 0$, $\forall q > \overline{q}_{\mu_2}(\alpha), \forall e_i \in [0, W]$, $e_j^* = W$ and all parameters in their domain, $e_i^* = 0$ is the optimum. Hence, $e_i^* = 0$ for $q \in [\overline{q}(\alpha), 1]$ is a local maximizer.

Since for $q \leq 0.5$, the best response of player i is always W, independent of the effort level of player j and when $q \geq 0.5$, the best response of player j is always W, independent of the effort level of player i, there exists an equilibrium $\forall a, q, W$. The equilibrium can be described by the following two best responses:

$$e_i^* = \begin{cases} W & \text{if } q \leq \overline{q}(\alpha) \\ 0 & \text{else} \end{cases}$$
$$e_j^* = \begin{cases} 0 & \text{if } q \leq (1 - \overline{q}(\alpha)) \\ W & \text{else} \end{cases}$$

□

B.2.3 Proof of Proposition 3

Proof of Proposition 3. We assume risk-neutral and spiteful players. Hence, the utilities from a successful settlement are the following: $U_i^{settlement} = s_i - \alpha s_j$ and $U_j^{settlement} = s_j - \alpha s_i$ and $d_i = U_i^{litigation}(e_i^*, e_j^*, q, \alpha)$ and $d_j = U_j^{litigation}(e_j^*, e_i^*, q, \alpha)$. We find the Nash bargaining solution by maximizing the function $f(s_i, s_j) = [(U_i^{settlement} - d_i)(U_j^{settlement} - d_j)]$ and solving the following optimization problem:

$$\max_{s_i, s_j}[(U_i^{settlement} - d_i)(U_j^{settlement} - d_j)]$$
$$\text{s.t. } s_i + s_j = W$$
$$s_i, s_j \geq 0$$
$$s_i, s_j \leq W$$

In the following, we use the KKT conditions to solve the optimization problem. The following KKT conditions have to be satisfied for a maximum. Note that $f(s_i, s_j)$ is concave, the inequality constraints convex, and the equality constraint affine, such that the KKT conditions are both necessary and sufficient. Let $h(s_i, s_j)$ denote the equality constraint and g_l the inequality constraints.

$$\begin{aligned}
\frac{\partial f(s_i,s_j)}{\partial s_i} - \lambda\frac{\partial h(s_i,s_j)}{\partial s_i} - \sum_{l=1}^{4}\mu_l\frac{\partial(g_l)}{\partial s_i} &= 0\\
\frac{\partial f(s_i,s_j)}{\partial s_j} - \lambda\frac{\partial h(s_i,s_j)}{\partial s_j} - \sum_{l=1}^{4}\mu_l\frac{\partial(g_l)}{\partial s_j} &= 0\\
h(s_i,s_j) = s_i + s_j - W &= 0\\
\mu_l(g_l(s_i,s_j)) = 0, \forall l &= 1,...,4\\
g_l(s_i,s_j) \le 0, \forall l &= 1,...,4\\
\mu_l \ge 0, \forall l &= 1,...,4
\end{aligned}$$

To determine the optimal requests, we first calculate the disagreement values. The equilibrium litigation behavior under the American fee-shifting rule is symmetric and is described by the following: $e_i^* = e_j^* = (1-q)qW(\alpha+1)$. Inserting these in the utility functions yields the expected utility of the respective players and hence the respective disagreement values: $d_i = U_i^{litigation}(e_i^*, e_j^*, q, \alpha) = (1 + (1-\alpha^2)q^2 - (2-\alpha^2+\alpha)q)W$ and $d_j = U_j^{litigation}(e_j^*, e_i^*, q, \alpha) = ((1-q)(q\alpha^2-\alpha)+q^2)W$.

We solve the optimization problem and find the following three points:

$$\begin{gathered}
(s_i^* = W, s_j^* = 0, \lambda = Wa^2q^2 - W\alpha^2q + W\alpha q - Wq^2 + 2Wq + \mu_3,\\
\mu_1 = 0, \mu_2 = -2W\alpha q + W\alpha - 2Wq - \mu_3, \mu_3 = \mu_3, \mu_4 = 0)
\end{gathered} \tag{B.6}$$

$$\begin{gathered}
(s_i^* = -W\alpha q + \frac{1}{2}W\alpha - qW + W, s_j^* = Waq - \frac{1}{2}W\alpha + qW,\\
\lambda = W\alpha^2q^2 - W\alpha^2q - Wq^2 + qW + \frac{1}{2}W\alpha, \mu_1 = 0, \mu_2 = 0, \mu_3 = 0, \mu_4 = 0)
\end{gathered} \tag{B.7}$$

$$\begin{gathered}
(s_i^* = 0, s_j^* = W, \lambda = W\alpha^2q^2 - W\alpha^2q + W\alpha q - Wq^2 + 2Wq - W - \mu_4,\\
\mu_1 = 2W\alpha q - W\alpha + 2Wq - 2W - \mu_4, \mu_2 = 0, \mu_3 = 0, \mu_4 = \mu_4)
\end{gathered} \tag{B.8}$$

The solutions are feasible within a specific region of q. First, we analyze the Point (B.6). Note that the solution allows for any value of $\mu_3 \geq 0$. Therefore, we put $\mu_3 = 0$ to get the threshold, where Point (B.6) is either just feasible or just not feasible anymore. We need to find the region where μ_2 is non-negative. Since $\frac{\partial \mu_2}{\partial q} = -2W\alpha - 2W \leq 0, \forall \alpha, W$ and $\mu_{2|q=0,\mu_3=0} > 0$, Point (B.6) is feasible until a certain threshold. We find this threshold by finding the root of μ_2. Hence, we put $-2W\alpha q + W\alpha - 2Wq = 0$ and solve it for q which yields: $q = \frac{1}{2}\frac{\alpha}{\alpha+1}$.

The analysis of the third point follows the same pattern. Point (B.8) is feasible after the threshold of $q = \frac{1}{2}\frac{\alpha+2}{\alpha+1}$.

Finally, we analyze Point (B.7). Note that for this point all conditions $g_l(s_i, s_j) \leq 0$ are met for $q \in (\frac{1}{2}\frac{\alpha}{\alpha+1}, \frac{1}{2}\frac{\alpha+2}{\alpha+1})$. Hence, the Nash bargaining solution in the American case can be described by the following function:

$$s_i^* = \begin{cases} W & \text{if } q \leq \frac{1}{2}\frac{\alpha}{\alpha+1} \\ W(a(\frac{1}{2} - q) + (1-q)) & \text{if } \frac{1}{2}\frac{\alpha}{\alpha+1} < q < \frac{1}{2}\frac{\alpha+2}{\alpha+1} \\ 0 & \text{if } q \geq \frac{1}{2}\frac{\alpha+2}{\alpha+1} \end{cases}$$

□

B.2.4 Proof of Proposition 4

Proof of Proposition 4. We find the Nash bargaining solution by maximizing the function

$$f(s_i, s_j) = [(U_i^{settlement} - d_i)(U_j^{settlement} - d_j)]$$

and solving the optimization problem described in Appendix B.2.3. The equilibrium litigation behavior is described by three zones:

1) $q < (1 - \overline{q}(\alpha))$ with $\overline{q}(\alpha) = \frac{1}{3}\frac{3\alpha+2}{\alpha+1} \in (\frac{2}{3}, \frac{5}{6})_{|\alpha\in(0,1)}$ from the litigation solution: In this region, $e_i^* = W$ and $e_j^* = 0$, which yield the following utilities: $d_i = U_i^{litigation} = (1+\alpha)W$ and $d_j = U_j^{litigation} = -(1+\alpha)W$. Solving the optimization problem yields the following KKT point:

$$(s_i^* = W, s_j^* = 0, \lambda = W\alpha^2 + W - \mu_2, \mu_1 = 0, \mu_2 = \mu_2, \mu_3 = W\alpha^2 + 2W\alpha + W - \mu_2, \mu_4 = 0)$$

This point is always feasible, because we can find a positive μ_2 for which μ_3 also becomes positive. Thus, in this region ($s_i^* = W$, $s_j^* = 0$).

2) $(1 - \overline{q}(\alpha)) < q < \overline{q}(\alpha)$: In this region: $e_i^* = e_j^* = W$ and the following utilities: $d_i = U_i^{litigation} = 3(1-q)(1+\alpha)W - 2W - \alpha W$ and $d_j = U_j^{litigation} = 3q(1+\alpha)W - 2W - \alpha W$.

Solving the optimization problem (9) yields the three following points:

$$(s_i^* = W, s_j^* = 0, \lambda = 3W\alpha^2 q + 6W\alpha q - 4W\alpha + 3Wq + \mu_3,$$
$$\mu_1 = 0, \mu_2 = -6W\alpha^2 q + 2Wa^2 - 12W\alpha q + 4W\alpha - 6Wq + 2W - \mu_3, \mu_3 = \mu_3, \mu_4 = 0) \tag{B.9}$$

$$(s_i^* = (2-3q)W, s_j^* = (3q-1)W, \lambda = W\alpha^2 - 2W\alpha + W, \mu_1 = 0, \mu_2 = 0, \mu_3 = 0, \mu_4 = 0) \tag{B.10}$$

$$(s_i^* = 0, s_j^* = W, \lambda = 3W\alpha^2 q - W\alpha^2 + 6W\alpha q - 6W\alpha + 3Wq - W - \mu_4,$$
$$\mu_1 = 6W\alpha^2 q - 4W\alpha^2 + 12W\alpha q - 8W\alpha + 6Wq - 4W - \mu_4, \mu_2 = 0, \mu_3 = 0, \mu_4 = \mu_4) \tag{B.11}$$

We check the KKT conditions for these three points. We start with Point (B.9). Note that the solution allows for any value of $\mu_3 \geq 0$. Therefore, we put $\mu_3 = 0$ to get the threshold where the point is either just feasible or just not feasible anymore. Since $\frac{\partial \mu_2}{\partial q} < 0, \forall \alpha, W$ and $\mu_{2|q=0,\mu_3=0} > 0$, we find the root of μ_2, which is the threshold $q = \frac{1}{3}$ until which Point (B.9) is feasible.

Now, we analyze Point (B.10). All conditions and the condition $g_l(s_i, s_j) \leq 0$ are met for $q \in [\frac{1}{3}, \frac{2}{3}]$.

Finally, we consider Point (B.11). First, we set $\mu_4 = 0$. Since $\frac{\partial \mu_1}{\partial q} > 0, \forall a, W$ and $\mu_{1|q=1,\mu_4=0=0} > 0$, we find the root of μ_1, which is the threshold $q = \frac{2}{3}$ from which on this point is feasible.

3) $q > \overline{q}(\alpha)$: In this region $e_i^* = 0$ and $e_j^* = W$, which yields disagreement values of $d_i = U_i^{litigation} = -(1+\alpha)W$ and $d_j = U_j^{litigation} = (1+\alpha)W$. Note that by symmetry of the Nash Bargaining solution, solving the optimization problem, and checking the KKT conditions yield the same but mirrored solution as in region 1): $(s_i^* = 0, s_j^* = W)$. Hence, we get the following solution:

$$s_i^* = \begin{cases} W & \text{if } q \leq \frac{1}{3} \\ (2-3q)W & \text{if } \frac{1}{3} < q < \frac{2}{3} \\ 0 & \text{if } q \geq \frac{2}{3} \end{cases}$$

□

B.2.5 Formal Derivations Hypotheses Litigation

Hypothesis 1.1: The average litigation expenditures of all merit levels q are higher under the English fee-shifting rule than under the American fee-shifting rule.

Proof. In order to compare the average litigation expenditures, it suffices to compare the aggregate expenditures. First, we calculate the aggregate litigation expenditures under the American rule:
$\int_0^1 e_j^{Am} dq = \int_0^1 q(1-q)W(\alpha+1)dq = \frac{1}{6}W(\alpha+1)$
The aggregate litigation expenditures under the English rule are as follows:
$\int_0^1 e_j^{Eng} dq = \int_0^{1-\overline{q}(\alpha)} 0dq + \int_{1-\overline{q}(\alpha)}^1 Wdq = \frac{1}{3}\frac{3\alpha+2}{\alpha+1}W$
Now, suppose the English expenditures are higher:

$$\frac{1}{3}\frac{3\alpha+2}{\alpha+1}W > \frac{1}{6}W(\alpha+1)$$
$$\iff 3\alpha+2 > \frac{1}{2}(\alpha+1)^2$$
$$\iff 4\alpha+3 > \alpha^2$$

which always holds true for $\alpha \in (0,1)$. □

Hypothesis 1.2: Under the American fee-shifting rule, average litigation expenditures are higher for more spiteful agents. This increase is driven by an increase at every merit level.

Proof. The litigation expenditures under the American fee-shifting rule are given by:

$$e^{*(\mathrm{Am})}(W,q,\alpha) = (1-q)\cdot q \cdot W \cdot (\alpha+1)$$

The derivative with respect to spite is as follows:

$$\frac{\partial e^{*(\mathrm{Am})}}{\partial \alpha} = (1-q)qW > 0 \forall \alpha \in (0,1) \text{ and } q \in (0,1)$$

As a consequence, spite always increases $e^{*(\mathrm{Am})}$ at every $\alpha \in (0,1)$ and $q \in (0,1)$. Hence, average expenditures over all merit levels q are also higher, independent of the distribution of α or q. □

Hypothesis 1.3: Under the English fee-shifting rule, average litigation expenditures over all merit levels q are higher for more spiteful agents. This increase is driven by an increase at low-merit levels while there is no increase at high-merit levels.

Proof. In order to compare average litigation expenditures, it suffices to compare aggregate litigation expenditures. Since $\frac{\partial(1-\overline{q}(\alpha))}{\partial a} = -\frac{1}{(3\alpha+3)^2} < 0$, the threshold $(1-\overline{q}(\alpha))$ to switch from spending 0 to W decreases for more spiteful litigants. Aggregated over all merit levels, litigation expenditures are the following:

$$\begin{aligned}\int_0^1 e_j^{*Eng} dq &= \int_0^{1-\overline{q}(\alpha)} 0 dq + \int_{1-\overline{q}(\alpha)}^1 W dq \\ &= W - \frac{1}{3(\alpha+1)} W\end{aligned}$$

Further, the aggregated expenditures increase in the spite level since

$$\frac{\partial(W - \frac{1}{3(\alpha+1)}W)}{\partial\alpha} = \frac{3}{(3\alpha+3)^2} W > 0$$

. □

Hypothesis 1.4: There is no difference in the increase of the average litigation expenditure over all q between the English and American fee-shifting rule for a non-spiteful ($\alpha \to 0$) compared to a fully spiteful ($\alpha \to 1$) agent.

Proof. First, we compute the aggregate litigation expenditures over all merits q (assuming a uniform distribution) and then compute the difference between a non-spiteful ($\alpha \to 0$) and fully spiteful ($\alpha \to 1$) agent. For the American rule, this is:

$$\int_0^1 e^{*(Am)} dq = \int_0^1 (1-q)qW(\alpha+1) dq = \frac{1}{6}W(\alpha+1)$$

$$\int_0^1 e^{*(Am)}(\alpha \to 0) dq = \frac{1}{6}W$$

$$\int_0^1 e^{*(Am)}(\alpha \to 1) dq = \frac{2}{6}W$$

$$\int_0^1 e^{*(Am)}(\alpha \to 1) dq - \int_0^1 e^{*(Am)}(\alpha \to 0) dq = \frac{1}{6}W$$

For the English rule this is:

$$\int_0^1 e_j^{*Eng} dq = \int_0^{1-\overline{q}(\alpha)} 0 dq + \int_{1-\overline{q}(\alpha)}^1 W dq = W - \frac{1}{3(\alpha+1)} W$$

$$\int_0^1 e^{*(Eng)}(\alpha \to 0) dq = \frac{2}{3}W$$

$$\int_0^1 e^{*(Eng)}(\alpha \to 1)dq = \frac{5}{6}W$$
$$\int_0^1 e^{*(Eng)}(\alpha \to 1)dq - \int_0^1 e^{*(Am)}(\alpha \to 0)dq = \frac{1}{6}W$$

□

B.2.6 Formal Derivations Hypotheses Settlement

Hypothesis 2.1: There is no difference in the average settlement requests over all merits q between the American and the English fee-shifting rule.

Proof. In order to compare the average settlement requests, it suffices to compare the aggregate requests. The aggregate settlement requests under the American rule are given by:

$$\begin{aligned}
\int_0^1 s_j^{Am} dq &= \int_0^{\frac{1}{2}\frac{\alpha}{\alpha+1}} 0dq + \int_{\frac{1}{2}\frac{\alpha}{\alpha+1}}^{\frac{1}{2}\frac{\alpha+2}{\alpha+1}} W(a(q-\frac{1}{2})+q)dq + \int_{\frac{1}{2}\frac{\alpha+2}{\alpha+1}}^{1} Wdq \\
&= [\frac{1}{2}Wq^2(\alpha+1) - \frac{1}{2}W\alpha q]\Big|_{\frac{1}{2}\frac{\alpha}{\alpha+1}}^{\frac{1}{2}\frac{\alpha+2}{\alpha+1}} + [Wq]\Big|_{\frac{1}{2}\frac{\alpha+2}{\alpha+1}}^{1} \\
&= [\frac{1}{2}W(\frac{1}{2}\frac{\alpha+2}{\alpha+1})^2(\alpha+1) - \frac{1}{2}W\alpha\frac{1}{2}\frac{\alpha+2}{\alpha+1}] - [\frac{1}{2}W(\frac{1}{2}\frac{\alpha}{\alpha+1})^2(\alpha+1) - \frac{1}{2}W\alpha\frac{1}{2}\frac{\alpha}{\alpha+1}] \\
&\quad + W - \frac{1}{2}\frac{\alpha+2}{\alpha+1}W \\
&= \frac{1}{8}W\frac{\alpha^2+4\alpha+4}{\alpha+1} - \frac{1}{4}W\frac{\alpha^2+2\alpha}{\alpha+1} - \frac{1}{8}W\frac{\alpha^2}{\alpha+1} + \frac{1}{4}W\frac{\alpha^2}{\alpha+1} + W - \frac{1}{2}W\frac{\alpha+2}{\alpha+1} \\
&= \frac{1}{8}W\frac{4\alpha+4}{\alpha+1} - \frac{1}{4}W\frac{2\alpha}{\alpha+1} + W - \frac{1}{2}\frac{\alpha+2}{\alpha+1} \\
&= \frac{1}{4}W\frac{2}{\alpha+1} + W - \frac{1}{4}W\frac{2\alpha+4}{\alpha+1} \\
&= W - \frac{1}{4}W\frac{2\alpha+2}{\alpha+1} \\
&= \frac{1}{2}W
\end{aligned}$$

The aggregate settlement requests under the English rule are given by:

$$\begin{aligned}
\int_0^1 s_j^{Eng} dq &= \int_0^{\frac{1}{3}} 0dq + \int_{\frac{1}{3}}^{\frac{2}{3}} (3q-1)Wdq + \int_{\frac{2}{3}}^{1} Wdq \\
&= [\frac{3}{2}q^2W - qW]\Big|_{\frac{1}{3}}^{\frac{2}{3}} + [Wq]\Big|_{\frac{2}{3}}^{1} \\
&= \frac{2}{3}W - \frac{2}{3}W - \frac{1}{6}W + \frac{1}{3}W + W - \frac{2}{3}W
\end{aligned}$$

$$= \frac{1}{2}W$$

□

Hypothesis 2.2: Under the American fee-shifting rule, the average settlement requests of low merits ($q < 0.5$) are lower for more spiteful agents, while for high merits ($q > 0.5$) they are higher.

Proof. Note that

$$s_j^{*Am} = \begin{cases} 0 & \text{if } q \leq \frac{1}{2}\frac{\alpha}{\alpha+1} \\ W(\alpha(q-\frac{1}{2})+q) & \text{if } \frac{1}{2}\frac{\alpha}{\alpha+1} < q < \frac{1}{2}\frac{\alpha+2}{\alpha+1} \\ W & \text{if } q \geq \frac{1}{2}\frac{\alpha+2}{\alpha+1} \end{cases}$$

This implies that for $q < \frac{1}{2}$, $s_j^* \in \{0, W(\alpha(q-\frac{1}{2})+q)\}$ and for $q > \frac{1}{2}$, $s_j^* \in \{W(\alpha(q-\frac{1}{2})+q), W\}$

Since $\frac{\partial \frac{1}{2}\frac{\alpha}{\alpha+1}}{\partial \alpha} = \frac{1}{2}\frac{1}{(\alpha+1)^2} > 0$, the threshold to switch from requesting 0 to $W(\alpha(q-\frac{1}{2})+q)$ increases for more spiteful agents for $q < 0.5$. Additionally, requests are lower after this threshold since $\frac{\partial W(\alpha(q-\frac{1}{2})+q)}{\partial \alpha} = W(q-\frac{1}{2}) < 0$ for $q < 0.5$. Hence, more spiteful agents, on average over low merits $q < 0.5$, request less for $q < 0.5$.

Since $\frac{\partial \frac{1}{2}\frac{\alpha+2}{\alpha+1}}{\partial \alpha} = -\frac{1}{2}\frac{1}{(\alpha+1)^2} < 0$, the threshold to switch from requesting $W(\alpha(q-\frac{1}{2})+q)$ to W decreases for more spiteful agents. Additionally, requests are higher before that threshold since $\frac{\partial W(\alpha(q-\frac{1}{2})+q)}{\partial \alpha} = W(q-\frac{1}{2}) > 0$ for $q > 0.5$. Hence, for $q > 0.5$, more spiteful agents, on average over high merits $q > 0.5$, request more. □

Hypothesis 2.3 Settlement requests under the English fee-shifting rule are the same for more spiteful and less spiteful agents.

Proof. This follows immediately from the equilibrium settlement requests, which are independent of the parameter α:

$$s_i^* = \begin{cases} W & \text{if } q \leq \frac{1}{3} \\ (2-3q)W & \text{if } \frac{1}{3} < q < \frac{2}{3} \\ 0 & \text{if } q \geq \frac{2}{3} \end{cases}$$

□

Finally, as in the efficient Nash-Demand game solution, all resources are always allocated without waste, there is no difference in the average influence of spite on the settlement requests over all merits q.

Hypothesis 2.4 There is no difference in the change of average settlement requests over all merits q between the English and American rule for non-spiteful compared to spiteful agents.

Proof. This follows immediately from the condition of the efficient Nash-Demand game solution, where $s_i + s_j = W$ for all q. Therefore, there is no difference in the average settlement requests over all merit q between the English and American rule independent of the spite level α. □

B.3 Main Regressions

B.3.1 American vs. English Fee-shifting Rule

To formally study the differences between the English and American rule, we use the following mixed-effects model with controls C_1 and C_2:[3]

$$\begin{aligned} D_{i,q} =& \beta_0 + \beta_1 q + \nu_i + \epsilon_{i,q} + C_M \\ C_1 =& 0 \\ C_2 =& \beta_{3Fee=Eng} + \beta_{4Fee=Eng} \times q \end{aligned} \tag{B.12}$$

where ν_i is a random effect for subject i, and $\epsilon_{i,q}$ is the residual. D is the dependent variable, which is either the litigation expenditures e or the settlement requests s. $_{Fee=Eng}$ denotes a dummy with value one if the fee-shifting rule is English and zero otherwise. Table B.1 shows the estimation results. Models (1) and (3) estimate the litigation expenditures under the American and English fee-shifting rules, respectively. Models (2) and (4) estimate the settlement requests under the American and English fee-shifting rule, respectively. Models (5) and (6) estimate Equation B.12 with Control C_2, i.e., the effect of fee-shifting on litigation expenditures and settlement requests, respectively.

It can be seen that both litigation expenditures and settlement requests are increasing significantly in merit q, giving support for the theory-derived functional form of the settlement request. For litigation expenditures, the observed behavior only follows tentatively the theoretical functional form. It can also be seen that under the English fee-shifting regime, both settlement requests and litigation expenditures are increasing significantly more compared to the American fee-shifting rule.

[3]We use a simple linear model assuming linearity in q. A look at Figure 2.3 validates the plausibility of this linearity assumption.

Table B.1: Mixed-effects regression of the litigation expenditures and settlement requests by fee-shifting rule as a function of q

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Comparison	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3.74***	4.07***	4.11***	3.99***	3.74***	4.07***
	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
Q	2.26***	1.97***	3.07***	2.23***	2.26***	1.97***
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
Eng					0.37***	−0.09*
					(0.06)	(0.05)
Q x Eng					0.81***	0.25***
					(0.10)	(0.08)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures following Equation B.12. Models (2), (4), and (6) estimate the settlement requests following Equation B.12. Models (1), (2), (3) and (4) estimate Equation B.12 with C_1. Models (5) and (6) estimate Equation B.12 with C_2. Standard errors are shown in parentheses. NumQ indicates the merit of the case while Eng denotes a dummy with value one for the English rule and zero for the American rule. $^{+}p<0.1$;$^{*}p<0.05$;$^{**}p<0.01$;$^{***}p<0.001$

B.3.2 Spiteful Preferences

Table B.2 reports on the mixed-effects regression for the aggregate legal expenditures and settlement requests by measure of social preference. Table B.3 reports on the mixed-effects regression for the aggregate legal expenditures and settlement requests by a continuous measure of social preference.

Table B.2: Mixed-effects regression of the litigation expenditures and settlement requests by measure of social-preferences

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
Panel A: Above/ below median spiteful subjects (via Spite-Task)						
Constant	4.24***	4.62***	5.50***	4.70***	5.50***	4.70***
	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)
MedianSpite	1.31***	0.93***	0.31***	0.85***	0.31***	0.85***
	(0.12)	(0.10)	(0.12)	(0.10)	(0.11)	(0.10)
American					−1.26***	−0.08*
					(0.06)	(0.04)
American:MedianSpite					1.00***	0.08
					(0.08)	(0.06)
Panel B: Above/ below median spiteful subjects (via Spite-Questionnaire)						
Constant	4.40***	4.58***	5.39***	4.64***	5.39***	4.64***
	(0.08)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)
MedianSpiteQ	0.95***	0.98***	0.53***	0.94***	0.53***	0.94***
	(0.12)	(0.10)	(0.12)	(0.10)	(0.11)	(0.10)
American					−0.99***	−0.06
					(0.06)	(0.04)
American:MedianSpiteQ					0.43***	0.04
					(0.08)	(0.06)
Panel C: Above/ below median prosocial subjects (via SVO-Measure)						
Constant	4.83***	5.15***	5.73***	5.17***	5.73***	5.17***
	(0.09)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)
MedianSVO	0.07	−0.18*	−0.16	−0.14	−0.16	−0.14
	(0.12)	(0.10)	(0.12)	(0.10)	(0.11)	(0.10)
American					−0.90***	−0.02
					(0.06)	(0.04)
American:MedianSVO					0.24***	−0.04
					(0.08)	(0.06)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. American denotes a dummy with value one for the American rule and zero for the English rule. MedianSpite/MedianSpiteQ/MedianSVO denotes a dummy with value one if the subject displays above-median preferences in the respective measure of social preferences. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table B.3: Mixed-effects regression of the litigation expenditures and settlement requests by a continuous measure of social-preferences

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
Panel A: Spiteful subjects (via Spite-Task)						
Constant	4.87***	5.06***	5.64***	5.10***	5.64***	5.10***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
ContinuousSpite	0.88***	0.69***	0.42***	0.66***	0.42***	0.66***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
American					−0.78***	−0.04
					(0.04)	(0.03)
American:ContinuousSpite					0.46***	0.03
					(0.04)	(0.03)
Panel B: Spiteful subjects (via Spite-Questionnaire)						
Constant	4.87***	5.06***	5.64***	5.10***	5.64***	5.10***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
ContinuousSpiteQ	0.87***	0.85***	0.54***	0.82***	0.54***	0.82***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
American					−0.78***	−0.04
					(0.04)	(0.03)
American:ContinuousSpiteQ					0.33***	0.03
					(0.04)	(0.03)
Panel C: Prosocial subjects (via SVO-Measure)						
Constant	4.87***	5.06***	5.64***	5.10***	5.64***	5.10***
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
ContinuousSVOMeasure	−0.005	−0.12**	−0.06	−0.10**	−0.06	−0.10**
	(0.06)	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)
American					−0.78***	−0.04
					(0.04)	(0.03)
American:ContinuousSVOMeasure					0.06	−0.02
					(0.04)	(0.03)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. American denotes a dummy with value one for the American rule and zero for the English rule. ContinuousSpite/ContinuousSpiteQ/ContinuousSVOMeasure denote the z-scored preferences in the respective measure of social preferences. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

B.3.3 The Effect of Spite as a Function of Merit

To study the effect of spite as a function of merit more formally, we use the following mixed-effects model with controls C_1, C_2, and C_3:

$$
\begin{aligned}
D_{i,q} =& \beta_0 + \beta_{1spite>median} + \nu_i + \epsilon_{i,q} + C_M \\
C_1 =& 0 \\
C_2 =& \beta_3 q + \beta_{4spite>median} \times q \\
C_3 =& C_2 + \beta_{5Fee=Am} + \beta_{6Fee=Am} \times q + \beta_{7Fee=Am} \times_{spite>median} \\
& + \beta_{8Fee=Am} \times q \times_{spite>median}
\end{aligned}
\tag{B.13}
$$

where ν_i is a random effect for subject i, and $\epsilon_{i,q}$ is the residual. D is the dependent variable, which is either the litigation expenditure e or the settlement request s. $_{Fee=Am}$ denotes a dummy with value one if the fee-shifting rule is American and zero otherwise. $_{spite>median}$ denotes a dummy with value one if the subject is more spiteful, i.e., if the subject scored higher than the median in the spite measurement. Table B.2 shows the estimation for litigation expenditures and settlement requests of Equation B.13 with control C_1, i.e, the effect of more spiteful vs. less spiteful subjects. Table B.4 shows the estimation of Equation B.13 with control C_2 and C_3, i.e the effect of more spiteful vs. less spiteful subjects as a function of merit q. Furthermore, Table B.5 replicates the result by using continuous measures of social preferences.

Table B.4: Mixed-effects regression of the litigation expenditures and settlement requests by measure of social-preferences as a function of q

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Above/ below median spiteful subjects (via Spite-Task)						
Constant	2.90***	3.33***	3.56***	3.22***	3.56***	3.22***
	(0.09)	(0.08)	(0.10)	(0.08)	(0.09)	(0.08)
NumQ	2.69***	2.57***	3.87***	2.95***	3.87***	2.95***
	(0.09)	(0.07)	(0.10)	(0.07)	(0.10)	(0.07)
MedianSpite	1.75***	1.55***	1.14***	1.60***	1.14***	1.60***
	(0.13)	(0.11)	(0.14)	(0.11)	(0.13)	(0.11)
NumQ:MedianSpite	−0.90***	−1.24***	−1.68***	−1.50***	−1.68***	−1.50***
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)
American					−0.67***	0.11*
					(0.09)	(0.06)
NumQ:American					−1.18***	−0.38***
					(0.14)	(0.10)
American:MedianSpite					0.61***	−0.05
					(0.13)	(0.09)
NumQ:American:MedianSpite					0.78***	0.26*
					(0.21)	(0.15)
Panel B: Above/ below median spiteful subjects (via Spite-Questionnaire)						
Constant	3.03***	3.38***	3.53***	3.32***	3.53***	3.32***
	(0.09)	(0.08)	(0.10)	(0.08)	(0.09)	(0.08)
NumQ	2.75***	2.41***	3.71***	2.66***	3.71***	2.66***
	(0.09)	(0.07)	(0.10)	(0.08)	(0.10)	(0.08)
MedianSpiteQ	1.46***	1.43***	1.18***	1.38***	1.18***	1.38***
	(0.14)	(0.11)	(0.14)	(0.11)	(0.13)	(0.11)
NumQ:MedianSpiteQ	−1.01***	−0.89***	−1.31***	−0.88***	−1.31***	−0.88***
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)
American					−0.51***	0.06
					(0.09)	(0.07)
NumQ:American					−0.96***	−0.25**
					(0.15)	(0.11)
American:MedianSpiteQ					0.27**	0.04
					(0.13)	(0.09)
NumQ:American:MedianSpiteQ					0.31	−0.01
					(0.21)	(0.15)
Panel C: Above/ below median prosocial subjects (via SVO-Measure)						
Constant	3.78***	4.12***	4.28***	4.03***	4.28***	4.03***
	(0.10)	(0.08)	(0.10)	(0.08)	(0.09)	(0.08)
NumQ	2.10***	2.07***	2.89***	2.28***	2.89***	2.28***
	(0.09)	(0.07)	(0.10)	(0.08)	(0.10)	(0.08)
MedianSVO	−0.09	−0.09	−0.34**	−0.09	−0.34**	−0.09
	(0.14)	(0.11)	(0.14)	(0.11)	(0.13)	(0.11)
NumQ:MedianSVO	0.32**	−0.19*	0.35**	−0.10	0.35**	−0.10
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)
American					−0.50***	0.08
					(0.09)	(0.07)
NumQ:American					−0.79***	−0.21*
					(0.15)	(0.11)
American:MedianSVO					0.25**	0.004
					(0.13)	(0.09)
NumQ:American:MedianSVO					−0.03	−0.09
					(0.21)	(0.15)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. NumQ indicates the merit of the case while American denotes a dummy with value one for the English rule and zero for the American rule. MedianSpite/MedianSpiteQ/MedianSVO denotes a dummy with value one if the subject displays above-median preferences in the respective measure of social preferences. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table B.5: Mixed-effects regression of the litigation expenditures and settlement requests by a continuous measure of social-preferences as a function of q

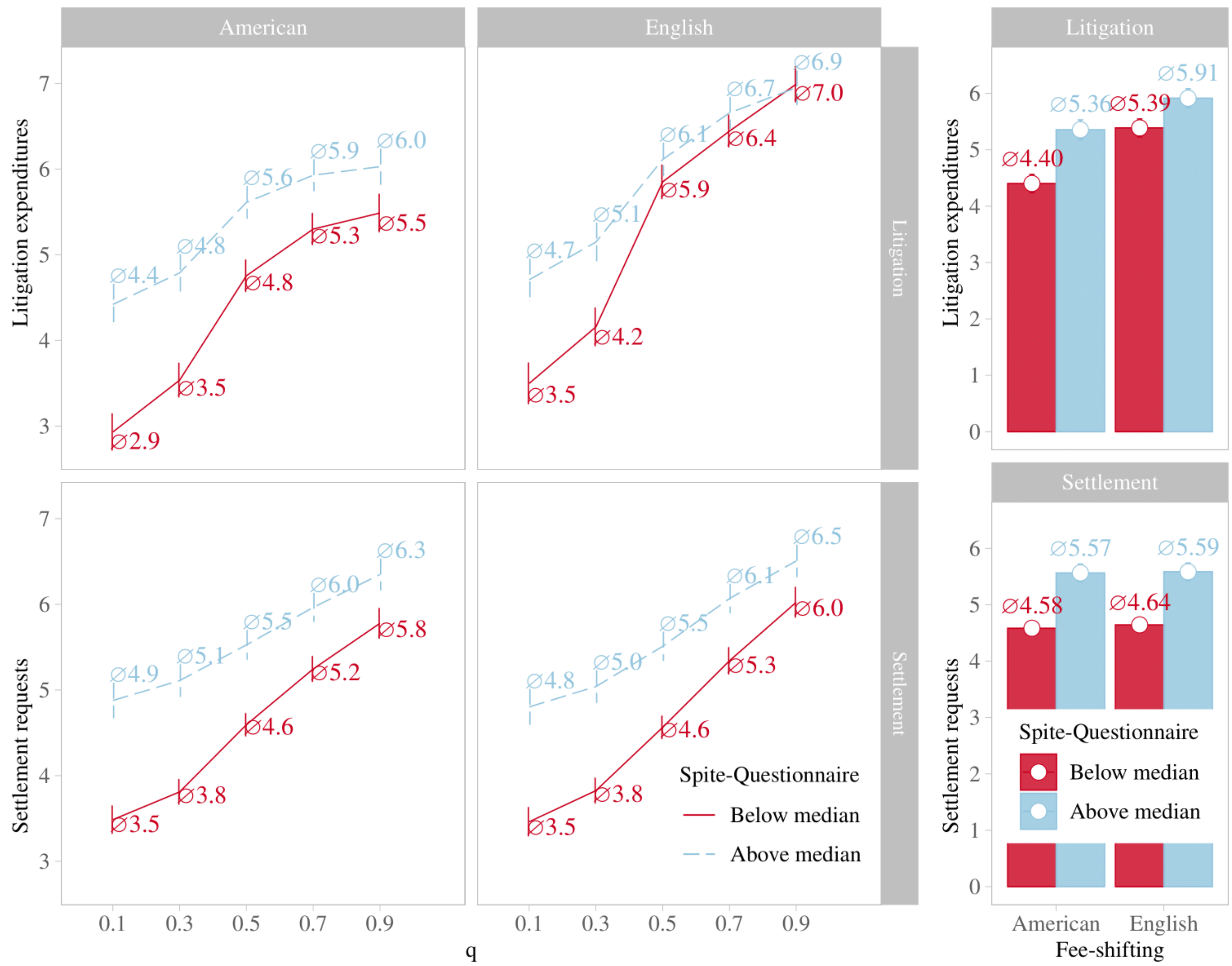
	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Above/ below median spiteful subjects (via Spite-Task)						
Constant	3.74***	4.07***	4.11***	3.99***	4.11***	3.99***
	(0.07)	(0.05)	(0.07)	(0.05)	(0.06)	(0.05)
NumQ	2.26***	1.97***	3.07***	2.23***	3.07***	2.23***
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
ContinuousSpite	1.18***	1.05***	0.95***	1.06***	0.95***	1.06***
	(0.07)	(0.05)	(0.07)	(0.05)	(0.06)	(0.05)
NumQ:ContinuousSpite	−0.60***	−0.72***	−1.06***	−0.81***	−1.06***	−0.81***
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
American					−0.37***	0.09*
					(0.06)	(0.05)
NumQ:American					−0.81***	−0.25***
					(0.10)	(0.08)
American:ContinuousSpite					0.23***	−0.01
					(0.06)	(0.05)
NumQ:American:ContinuousSpite					0.46***	0.09
					(0.10)	(0.08)
Panel B: Above/ below median spiteful subjects (via Spite-Questionnaire)						
Constant	3.74***	4.07***	4.11***	3.99***	4.11***	3.99***
	(0.07)	(0.05)	(0.07)	(0.05)	(0.06)	(0.05)
NumQ	2.26***	1.97***	3.07***	2.23***	3.07***	2.23***
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
ContinuousSpiteQ	1.24***	1.16***	1.04***	1.15***	1.04***	1.15***
	(0.07)	(0.05)	(0.07)	(0.05)	(0.06)	(0.05)
NumQ:ContinuousSpiteQ	−0.74***	−0.63***	−1.00***	−0.65***	−1.00***	−0.65***
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
American					−0.37***	0.09*
					(0.06)	(0.05)
NumQ:American					−0.81***	−0.25***
					(0.10)	(0.08)
American:ContinuousSpiteQ					0.20***	0.01
					(0.06)	(0.05)
NumQ:American:ContinuousSpiteQ					0.26**	0.02
					(0.10)	(0.08)
Panel C: Above/ below median prosocial subjects (via SVO-Measure)						
Constant	3.74***	4.07***	4.11***	3.99***	4.11***	3.99***
	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
NumQ	2.26***	1.97***	3.07***	2.23***	3.07***	2.23***
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
ContinuousSVOMeasure	−0.10	−0.11*	−0.18**	−0.12**	−0.18***	−0.12**
	(0.07)	(0.06)	(0.07)	(0.06)	(0.07)	(0.06)
NumQ:ContinuousSVOMeasure	0.19***	−0.02	0.23***	0.04	0.23***	0.04
	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.05)
American					−0.37***	0.09*
					(0.06)	(0.05)
NumQ:American					−0.81***	−0.25***
					(0.10)	(0.08)
American:ContinuousSVOMeasure					0.08	0.02
					(0.06)	(0.05)
NumQ:American:ContinuousSVOMeasure					−0.04	−0.07
					(0.10)	(0.08)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. NumQ indicates the merit of the case while American denotes a dummy with value one for the English rule and zero for the American rule. ContinuousSpite/ContinuousSpiteQ/ContinuousSVOMeasure denotes the z-scored preferences in the respective measure of social preferences. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

The econometric estimations roughly confirm our visual inspections. Litigation expenditures and settlement requests increase with merit. More spiteful subjects (measures by both measures of spite) start off with a substantially higher settlement request and substantially higher litigation expenditures. However, with increasing merit, the difference between more and less spiteful subjects in litigation expenditures and settlement requests decreases (as β_4 is significantly negative). Yet, the difference remains always positive under the American fee-shifting rule. Thus, we find further support for Hypothesis 1.2, stating that under the American fee-shifting rule, litigation expenditures are higher for more spiteful agents at every level of merit. Further, we see that more spiteful subjects request more than less spiteful subjects for all merit levels (under both fee-shifting rules). Thus, we find no support for the first part of Hypothesis 2.2 – stating that under the American fee-shifting rule, more spiteful subjects request less than less spiteful subjects for low-merit cases – no support for Hypothesis 2.3 – claiming no difference in settlement requests under the English fee-shifting rule between more and less spiteful subjects – and some support for the second part of Hypothesis 2.2 – stating that under the American fee-shifting rule, more spiteful subjects request more than less spiteful subjects for high-merit cases.

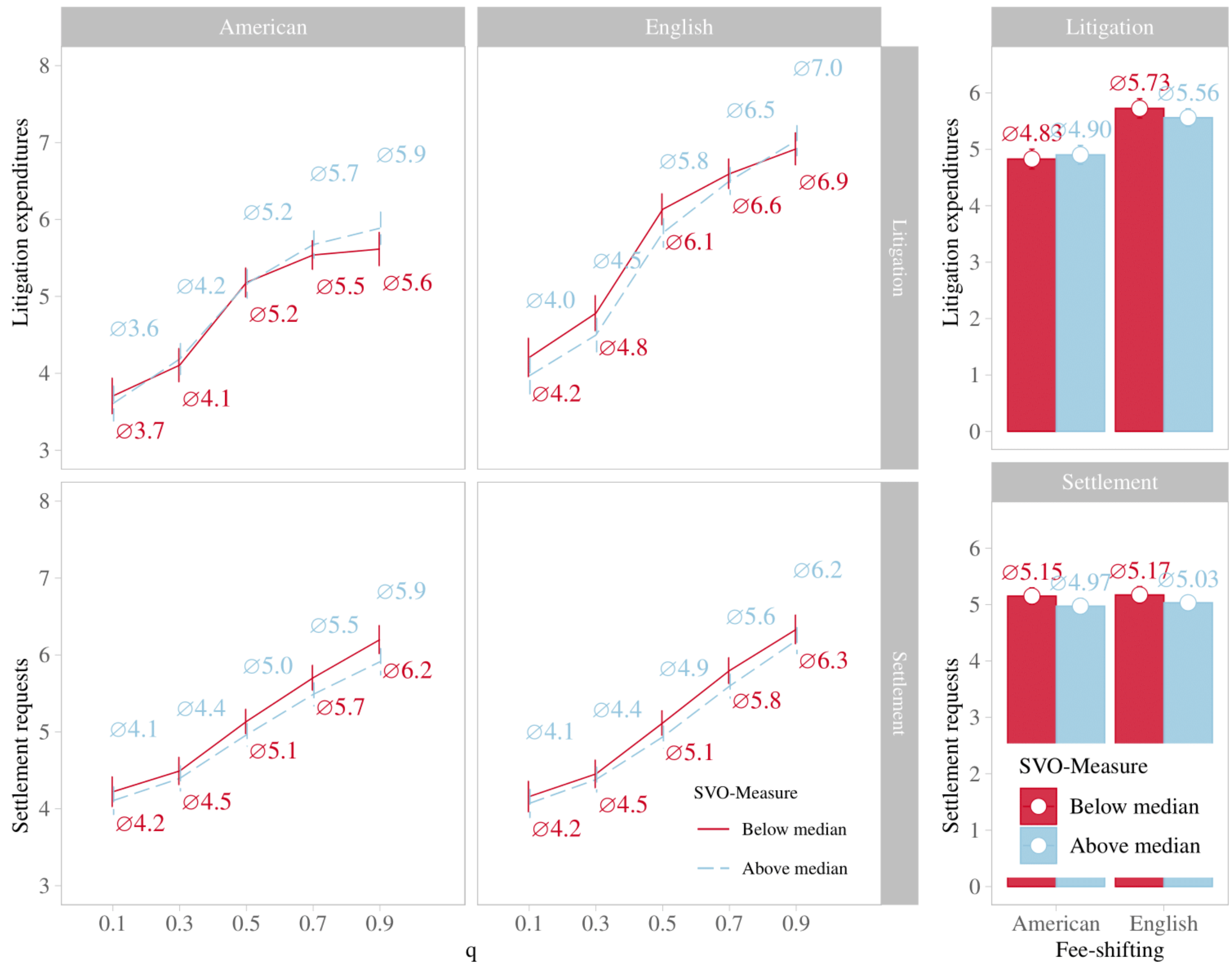
We further find that the difference between more and less spiteful subjects is more pronounced under the American fee-shifting rule compared to the English fee-shifting rule (as β_7 is significantly positive and β_8 is also positive). We find only little evidence for such a difference in the fee-shifting rules for the settlement requests between more and less spiteful subjects. The following figures illustrate the effect of spite as a function of merit and fee-shifting rule for the Spite-Questionnaire and the SVO-Measure.

Figure B.1: Litigation effort and settlement requests under the American and English fee-shifting rule as a function of q for more and less spiteful subjects (classified via the Spite-Questionnaire)



Note: The figures to the left depict the litigation expenditures and settlement requests by fee-shifting rule for more and less spiteful subjects as a function of q, while the figures to the right show the aggregates. The panels on the top illustrate the litigation effort, while the panels on the bottom illustrate the settlement requests. Red solid lines depict the behavior of less spiteful subjects (i.e., subjects with below-median spite scores on the Spite-Questionnaire), while blue dashed lines indicate the response of more spiteful subjects. The error bars indicate the 95% confidence intervals.

Figure B.2: Litigation effort and settlement requests under the American and English fee-shifting rule as a function of q for more and less prosocial subjects



Note: The figures to the left depict the litigation expenditures and settlement requests by fee-shifting rule for more and less prosocial subjects as a function of q, while the figures to the right depict the aggregates. The panels on the top depict the litigation effort, while the panels on the bottom illustrate the settlement requests. Red solid lines depict the behavior of less prosocial subjects (i.e., subjects with below-median SVO-scores on the SVO-Measure), while blue dashed lines indicate the behavior of more prosocial subjects. The error bars indicate the 95% confidence intervals.

B.3.4 The Costs of Spite

Table B.6: Mixed-effects regression of the expected payoff by fee-shifting rule and spiteful preferences

	Dependent Variable:								
	Expected Payoff Matched with Whole Population			*Expected Payoff* Matched with Below Spite			*Expected Payoff* Matched with Above Spite		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	2.22^{***}	-1.96^{***}	-4.08^{***}	2.80^{***}	-1.20^{***}	-3.06^{***}	1.86^{***}	-2.32^{***}	-4.55^{***}
	(0.07)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
American	0.47^{***}	0.47^{***}	4.71^{***}	0.33^{***}	0.33^{***}	4.04^{***}	0.34^{***}	0.34^{***}	4.79^{***}
	(0.09)	(0.06)	(0.06)	(0.08)	(0.05)	(0.06)	(0.09)	(0.06)	(0.06)
MedianSpite	-0.92^{***}	-0.92^{***}	-0.92^{***}	-0.86^{***}	-0.86^{***}	-0.86^{***}	-0.88^{***}	-0.88^{***}	-0.88^{***}
	(0.10)	(0.08)	(0.08)	(0.09)	(0.08)	(0.08)	(0.10)	(0.08)	(0.08)
NumQ		8.35^{***}	12.60^{***}		8.00^{***}	11.71^{***}		8.38^{***}	12.82^{***}
		(0.06)	(0.07)		(0.06)	(0.06)		(0.06)	(0.07)
American:MedianSpite	-0.15	-0.15^{+}	-0.15^{*}	-0.22^{+}	-0.22^{**}	-0.22^{***}	-0.18	-0.18^{*}	-0.18^{**}
	(0.13)	(0.08)	(0.07)	(0.12)	(0.08)	(0.06)	(0.13)	(0.08)	(0.07)
American:NumQ			-8.48^{***}			-7.43^{***}			-8.89^{***}
			(0.09)			(0.09)			(0.09)
Observations	16,350	16,350	16,350	16,350	16,350	16,350	16,350	16,350	16,350

Models (1), (2), and (3) estimate the expected payoff when being matched with the whole population of the experiment. Models (4), (5), and (6) estimate the expected payoff when being matched only with below median-spite subjects. Models (7), (8), and (9) estimate the expected payoff when being matched only with above median spite subjects. NumQ indicates the merit of the case while American denotes a dummy with value one for the English rule and zero for the American rule. MedianSpite denotes a dummy with value one if the subject displays above-median spite. All models estimate the interaction between the fee-shifting rules and above-median spite. Models (2), (5), and (7) additionally control for the merit of the case, while models (3), (6), (8) further control for the interaction of the merit and the fee-shifting rule. Standard errors are shown in parentheses. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table B.7: Mixed-effects regression of the expected payoff by fee-shifting rule and being matched with above or below spite opponents

	Dependent Variable:								
	Expected Payoff Whole Population			*Expected Payoff* Less Spiteful Subjects			*Expected Payoff* More Spiteful Subjects		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Constant	2.49***	−1.70***	−3.67***	2.90***	−0.79***	−2.59***	2.04***	−2.68***	−4.87***
	(0.07)	(0.06)	(0.06)	(0.08)	(0.07)	(0.07)	(0.11)	(0.09)	(0.09)
American	0.16+	0.13*	4.19***	0.25*	0.25***	4.04***	0.08	−0.01	4.40***
	(0.09)	(0.06)	(0.07)	(0.11)	(0.07)	(0.08)	(0.14)	(0.09)	(0.10)
MatchedSpiteAbove	−0.87***	−0.89***	−0.90***	−0.92***	−0.90***	−0.90***	−0.81***	−0.89***	−0.93***
	(0.09)	(0.06)	(0.05)	(0.10)	(0.07)	(0.06)	(0.14)	(0.09)	(0.07)
NumQ		8.25***	12.14***		7.22***	10.75***		9.37***	13.71***
		(0.06)	(0.06)		(0.07)	(0.08)		(0.09)	(0.10)
American:MatchedSpiteAbove	−0.11	0.08	0.02	−0.04	0.08	0.002	−0.20	0.09	0.06
	(0.12)	(0.08)	(0.07)	(0.15)	(0.10)	(0.08)	(0.20)	(0.12)	(0.10)
American:NumQ			−7.99***			−7.41***			−8.72***
			(0.09)			(0.11)			(0.14)
Observations	16,350	16,350	16,350	8,502	8,502	8,502	7,848	7,848	7,848

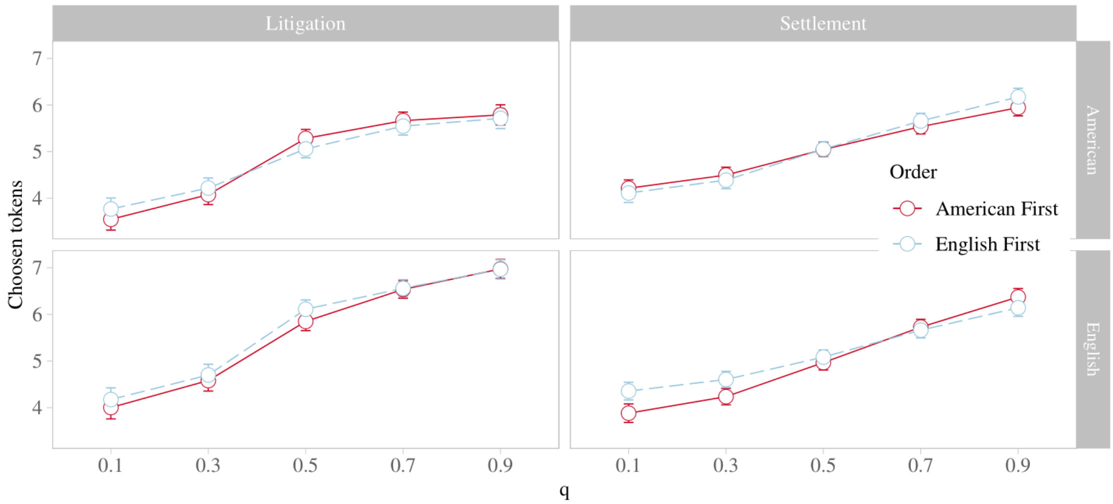
Models (1), (2), and (3) estimate the expected payoff for the whole population of the experiment. Models (4), (5), and (6) estimate the expected payoff only for below-median spite subjects. Models (7), (8), and (9) estimate the expected payoff only for above-median spite subjects. NumQ indicates the merit of the case while American denotes a dummy with value one for the English rule and zero for the American rule. MatchedSpiteAbove denotes a dummy with value one if the subject is matched only with above-median spite subjects. All models estimate the interaction between the fee-shifting rules and above-median spite. Models (2), (5), and (7) additionally control for the merit of the case, while models (3), (6), (8) further control for the interaction of the merit and the fee-shifting rule. Standard errors are shown in parentheses. $^{+}p<0.1$;$^{*}p<0.05$;$^{**}p<0.01$;$^{***}p<0.001$

B.4 Further Regressions

B.4.1 Order Effects

In this subsection, we show that there is no order effect of the fee-shifting rule. In particular, Figure B.3 shows the litigation expenditures and settlement requests as a function of the merit under the American and the English fee-shifting rule both if the American fee-shifting rule is played first and if the English fee-shifting rule is played first. Table B.8 provides the corresponding regression. No order effect can be found.

Figure B.3: Order effects on litigation expenditures and settlement requests under both fee-shifting rules



Note: The panels on the left depict the litigation effort, while the panels on the right illustrate the settlement requests. The panels on top show the behavior under the American rule, while the panels on the bottom show the behavior under the English rule. Red solid lines depict the behavior if the American rule was played first, while blue dashed lines indicate the response if the English rule was played first in each panel.

Table B.8: Mixed-effects regression of order effects on litigation expenditures and settlement requests under both fee-shifting rules

	Dependent Variable:			
	Litigation/ Settlement			
	American		English	
	(1)	(2)	(3)	(4)
Constant	4.87***	5.05***	5.59***	5.04***
	(0.09)	(0.07)	(0.08)	(0.07)
EnglishFirst	−0.01	0.03	0.11	0.13
	(0.12)	(0.10)	(0.12)	(0.10)
Litigation	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175
Note:	$^{+}$p<0.1;*p<0.05;**p<0.01;***p<0.001			

Models (1) and (3) estimate the litigation expenditures. Models (2) and (4) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Standard errors are shown in parentheses. EnglishFirst denotes a dummy with value one if the English rule was played first and zero if the American rule was played first. $^{+}$*p<0.1;***p<0.05;****p<0.01;*****p<0.001*

B.4.2 Additional Controls

In this section, we employ a robustness check by running models that include controls (age, gender, and educational attainment) and risk preferences (see Section B.4.4 for a discussion of the measure). Table B.9 shows the results of these regressions. We observe that all main findings remain robust to these alternative model specifications.

Table B.9: Main mixed-effects regression with controls

	Dependent Variable:											
	Litigation/ Settlement											
	American		English		Both		American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Above/ below median spiteful subjects (via Spite-Task)												
Constant	2.72***	3.59***	4.09***	3.60***	3.74***	3.54***	1.67***	2.48***	3.21***	2.66***	2.79***	2.54***
	(0.23)	(0.19)	(0.23)	(0.19)	(0.20)	(0.18)	(0.34)	(0.30)	(0.35)	(0.29)	(0.31)	(0.28)
NumQ	2.69***	2.57***	3.87***	2.95***	3.87***	2.95***	2.85***	2.61***	3.97***	2.87***	3.97***	2.87***
	(0.09)	(0.07)	(0.10)	(0.07)	(0.10)	(0.07)	(0.12)	(0.10)	(0.14)	(0.11)	(0.14)	(0.11)
MedianSpite	1.76***	1.55***	1.14***	1.59***	1.14***	1.59***	2.33***	2.02***	1.75***	1.94***	1.75***	1.94***
	(0.13)	(0.11)	(0.14)	(0.11)	(0.13)	(0.11)	(0.18)	(0.16)	(0.19)	(0.16)	(0.18)	(0.16)
American					−0.67***	0.11*					−0.70***	0.06
					(0.09)	(0.06)					(0.12)	(0.09)
NumQ:American					−1.18***	−0.38***					−1.12***	−0.27*
					(0.14)	(0.10)					(0.20)	(0.15)
bretrisk							0.02***	0.02***	0.02***	0.02***	0.02***	0.02***
							(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
NumQ:MedianSpite	−0.90***	−1.24***	−1.68***	−1.50***	−1.68***	−1.50***	−1.59***	−1.68***	−2.42***	−1.77***	−2.42***	−1.77***
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)	(0.16)	(0.14)	(0.19)	(0.15)	(0.19)	(0.15)
American:MedianSpite					0.61***	−0.05					0.58***	0.08
					(0.13)	(0.09)					(0.17)	(0.13)
NumQ:American:MedianSpite					0.78***	0.26*					0.83***	0.09
					(0.21)	(0.15)					(0.28)	(0.21)
Panel B: Above/ below median spiteful subjects (via Spite-Questionnaire)												
Constant	2.57***	3.39***	4.01***	3.47***	3.55***	3.39***	1.51***	2.32***	2.99***	2.42***	2.54***	2.30***
	(0.25)	(0.20)	(0.24)	(0.20)	(0.22)	(0.19)	(0.35)	(0.30)	(0.36)	(0.30)	(0.32)	(0.29)
NumQ	2.75***	2.41***	3.71***	2.66***	3.71***	2.66***	2.95***	2.46***	4.01***	2.80***	4.01***	2.80***
	(0.09)	(0.07)	(0.10)	(0.08)	(0.10)	(0.08)	(0.12)	(0.11)	(0.14)	(0.11)	(0.15)	(0.11)
MedianSpiteQ	1.47***	1.37***	1.03***	1.30***	1.11***	1.31***	2.29***	1.98***	1.93***	2.05***	1.94***	2.06***
	(0.14)	(0.11)	(0.14)	(0.11)	(0.14)	(0.11)	(0.19)	(0.16)	(0.20)	(0.16)	(0.19)	(0.16)
American					−0.51***	0.06					−0.57***	0.15
					(0.09)	(0.07)					(0.13)	(0.09)
NumQ:American					−0.96***	−0.25**					−1.07***	−0.33**
					(0.15)	(0.11)					(0.21)	(0.15)
bretrisk							0.02***	0.02***	0.02***	0.02***	0.02***	0.02***
							(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
NumQ:MedianSpiteQ	−1.01***	−0.89***	−1.31***	−0.88***	−1.31***	−0.88***	−1.71***	−1.35***	−2.40***	−1.56***	−2.40***	−1.56***
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)	(0.16)	(0.14)	(0.19)	(0.15)	(0.20)	(0.15)
American:MedianSpiteQ					0.27**	0.04					0.33*	−0.08
					(0.13)	(0.09)					(0.17)	(0.13)
NumQ:American:MedianSpiteQ					0.31	−0.01					0.70**	0.20
					(0.21)	(0.15)					(0.28)	(0.21)
Panel C: Above/ below median prosocial subjects (via SVO-Measure)												
Constant	3.71***	4.40***	4.77***	4.43***	4.49***	4.37***	2.94***	3.54***	4.32***	3.66***	3.95***	3.54***
	(0.23)	(0.19)	(0.22)	(0.19)	(0.20)	(0.18)	(0.35)	(0.30)	(0.34)	(0.30)	(0.31)	(0.29)
NumQ	2.10***	2.07***	2.89***	2.28***	2.89***	2.28***	1.86***	1.82***	2.52***	2.07***	2.52***	2.07***
	(0.09)	(0.07)	(0.10)	(0.08)	(0.10)	(0.08)	(0.12)	(0.10)	(0.13)	(0.10)	(0.14)	(0.10)
MedianSVO	−0.05	−0.02	−0.25*	−0.02	−0.28**	−0.02	0.15	0.13	−0.38**	0.16	−0.37**	0.15
	(0.14)	(0.11)	(0.14)	(0.11)	(0.13)	(0.11)	(0.19)	(0.16)	(0.19)	(0.16)	(0.19)	(0.16)
American					−0.50***	0.08					−0.64***	0.11
					(0.09)	(0.07)					(0.12)	(0.09)
NumQ:American					−0.79***	−0.21*					−0.66***	−0.25*
					(0.15)	(0.11)					(0.20)	(0.15)
bretrisk							0.02***	0.02***	0.02***	0.02***	0.02***	0.02***
							(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
NumQ:MedianSVO	0.32**	−0.19*	0.35**	−0.10	0.35**	−0.10	0.30*	−0.21	0.35*	−0.26*	0.35*	−0.26*
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)	(0.17)	(0.14)	(0.19)	(0.15)	(0.20)	(0.15)
American:MedianSVO					0.25**	0.004					0.52***	−0.02
					(0.13)	(0.09)					(0.17)	(0.13)
NumQ:American:MedianSVO					−0.03	−0.09					−0.05	0.05
					(0.21)	(0.15)					(0.28)	(0.21)
Litigation	✓	×	✓	×	✓	×	✓	×	✓	×	✓	×
Controls	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	8,175	8,175	8,175	8,175	16,350	16,350	4,110	4,110	4,110	4,110	8,220	8,220

Models (1), (3), (5), (7), (9), and (11) estimate the litigation expenditures. Models (2), (4), (6), (8), (10), and (12) estimate the settlement requests. Models (1), (2), (7), and (8) estimate behavior under the American fee-shifting rule. Models (3), (4), (9), and (10) estimate behavior under the English fee-shifting rule. Models (5), (6), (11), and (12) additionally estimate the interaction between both fee-shifting rules. Models (7), (8), (9), (10), (11), and (12) additionally control for risk-aversion (bretrisk). Standard errors are shown in parentheses. NumQ indicates the merit of the case while American denotes a dummy with value one for the American rule and zero for the English rule. MedianSpite/MedianSpiteQ/MedianSVO denotes a dummy with value one if the subject displays above-median preferences in the respective measure of social preferences. All models account for age, gender, and educational attainment. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

B.4.3 Wave Effects

In this section, we analyze whether the behavior between the two waves significantly differs. Table B.12 and Table B.13 report on the mixed-effects regressions comparing the two waves. We see a general tendency to request more in the second wave in the settlement stage. We also see that litigation expenditures are higher in the second wave. Further, we find some small effects in the function of Q between the two waves. More importantly, however, we find no interaction effect between the wave and the fee-shifting rule.

Table B.10: Mixed-effects regression of the litigation expenditures and settlement requests for the first and second wave

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	4.59***	4.75***	5.41***	4.82***	5.41***	4.82***
	(0.09)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)
American					−0.82***	−0.07*
					(0.06)	(0.04)
Wave	0.55***	0.62***	0.46***	0.56***	0.46***	0.56***
	(0.12)	(0.10)	(0.12)	(0.10)	(0.11)	(0.10)
American:Wave					0.09	0.07
					(0.08)	(0.06)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. American denotes a dummy with value one for the American rule and zero for the English rule. Wave denotes a dummy with value one if the behavior of the second wave is depicted and zero otherwise. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table B.11: Mixed-effects regression of the litigation expenditures and settlement requests as a function of q for the first and second wave

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3.33***	3.63***	3.69***	3.56***	3.69***	3.56***
	(0.10)	(0.08)	(0.10)	(0.08)	(0.09)	(0.08)
NumQ	2.51***	2.23***	3.45***	2.52***	3.45***	2.52***
	(0.09)	(0.07)	(0.10)	(0.08)	(0.11)	(0.08)
American					−0.35***	0.07
					(0.09)	(0.07)
Wave	0.80***	0.88***	0.84***	0.84***	0.84***	0.84***
	(0.14)	(0.11)	(0.14)	(0.11)	(0.13)	(0.11)
NumQ:American					−0.94***	−0.29***
					(0.15)	(0.11)
NumQ:Wave	−0.51***	−0.51***	−0.76***	−0.58***	−0.76***	−0.58***
	(0.12)	(0.10)	(0.14)	(0.11)	(0.15)	(0.11)
American:Wave					−0.04	0.03
					(0.13)	(0.09)
NumQ:American:Wave					0.25	0.07
					(0.21)	(0.15)
Litigation	✓	×	✓	×	✓	×
Observations	8,175	8,175	8,175	8,175	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. NumQ indicates the merit of the case while American denotes a dummy with value one for the American rule and zero for the English rule. Wave denotes a dummy with value one if the behavior of the second wave is depicted and zero otherwise. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

B.4.4 Risk Preferences

In the second wave, participants performed the bomb risk elicitation task (Crosetto and Filippin, 2013). In this task, subjects are presented with an interface that consists of 100 boxes. One of these boxes contains a bomb. Subjects are asked to choose how many boxes to select. If one of the chosen boxes contains the bomb, their earnings are zero. Otherwise, they earn 1 point for every box that they choose to open. In this task, a risk-neutral subject would choose 50 boxes. Higher values are indicative of risk-seeking preferences and lower values as risk-aversion. We use the number of boxes chosen by the participants as their preferences to take risks.

Here, we look at how litigation expenditures and settlement requests are related to risk preferences. As before, we employ median splits. We classify subjects as risk-seeking if their score is higher than the median risk score and non-risk-seeking (risk-averse) otherwise. Table B.12 and Table B.13 report on the mixed-effects regressions and Figure B.4 illustrates the results. Consistent with the literature, we see that higher levels of risk are related to higher litigation expenditures and settlement requests. However, we do not find any statistically significant differences in this relationship between the American and the English fee-shifting rule. We further find that the influence of risk decreases with increasing merit.

Table B.12: Mixed-effects regression of the litigation expenditures and settlement requests with median risk splits

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	4.62***	4.88***	5.41***	4.91***	5.41***	4.91***
	(0.12)	(0.11)	(0.12)	(0.10)	(0.11)	(0.10)
American					−0.79***	−0.03
					(0.07)	(0.06)
MedianRisk	1.06***	1.00***	0.96***	0.96***	0.96***	0.96***
	(0.17)	(0.15)	(0.17)	(0.15)	(0.16)	(0.15)
American:MedianRisk					0.11	0.04
					(0.11)	(0.08)
Litigation	✓	×	✓	×	✓	×
Observations	4,110	4,110	4,110	4,110	8,220	8,220

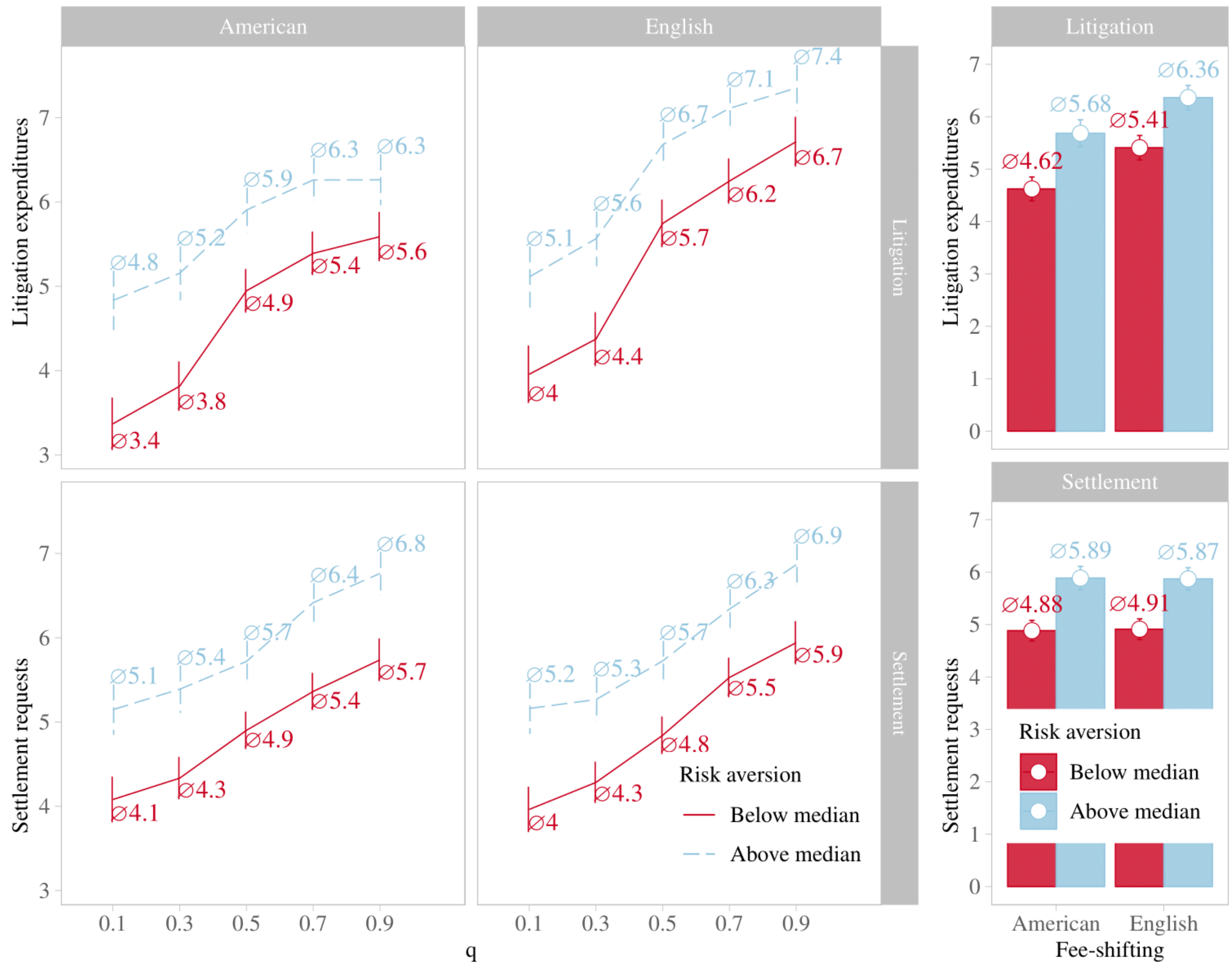
Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. American denotes a dummy with value one for the American rule and zero for the English rule. MedianRisk denotes a dummy with value one if the subject displays above-median risk-seeking preferences. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Table B.13: Mixed-effects regression of the litigation expenditures and settlement requests with median risk splits as a function of q

	Dependent Variable:					
	Litigation/ Settlement					
	American		English		Both	
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	3.42***	4.01***	3.93***	3.87***	3.93***	3.87***
	(0.13)	(0.12)	(0.13)	(0.12)	(0.13)	(0.12)
NumQ	2.41***	1.74***	2.96***	2.08***	2.96***	2.08***
	(0.12)	(0.10)	(0.13)	(0.10)	(0.14)	(0.10)
American					−0.51***	0.14
					(0.12)	(0.09)
MedianRisk	1.47***	1.02***	1.23***	1.10***	1.23***	1.10***
	(0.19)	(0.17)	(0.19)	(0.16)	(0.19)	(0.17)
NumQ:American					−0.55***	−0.34**
					(0.20)	(0.15)
NumQ:MedianRisk	−0.82***	−0.03	−0.55***	−0.29*	−0.55***	−0.29*
	(0.17)	(0.14)	(0.19)	(0.15)	(0.20)	(0.15)
American:MedianRisk					0.24	−0.08
					(0.17)	(0.13)
NumQ:American:MedianRisk					−0.27	0.25
					(0.28)	(0.21)
Litigation	✓	×	✓	×	✓	×
Observations	4,110	4,110	4,110	4,110	8,220	8,220

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate behavior under the American fee-shifting rule. Models (3) and (4) estimate under the English fee-shifting rule. Models (5) and (6) additionally estimate the interaction between both fee-shifting rules. Standard errors are shown in parentheses. NumQ indicates the merit of the case while American denotes a dummy with value one for the American rule and zero for the English rule. MedianRisk denotes a dummy with value one if the subject displays above-median risk-seeking preferences. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

Figure B.4: Litigation effort and settlement requests under the American and English fee-shifting rule as a function of q for risk-seeking and risk-averse subjects



Note: The figures to the left depict the litigation expenditures and settlement requests by fee-shifting rule for risk-seeking and risk-averse subjects as a function of q, while the figures to the right show the aggregates. The panels on the top illustrate the litigation effort, while the panels on the bottom illustrate the settlement requests. Red solid lines depict the behavior of risk-averse subjects (i.e., subjects with below-median scores in the Bomb-task), while blue dashed lines indicate the response of risk-seeking subjects (i.e., subjects with above-median scores in the Bomb-task). The error bars indicate the 95% confidence intervals.

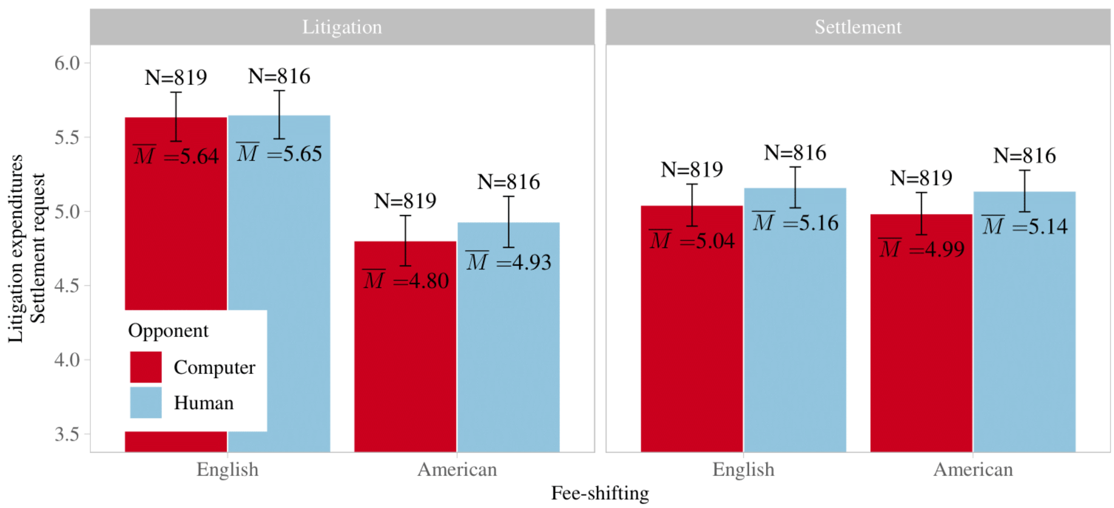
B.5 Causality

In this paper, we provide correlational evidence for a relationship between subjects with spiteful preferences and higher litigation expenditures and settlement requests. However, social-preferences are not exogenously assigned to subjects. To deal with this issue, we tried to manipulate spite exogenously.

To do so, we conducted the following additional treatments: In the baseline treatment, subjects were competing with a fellow participant. As we are not aware of any way to directly manipulate spiteful preferences, we decided to exclude social preferences altogether. To exclude social preferences, we matched subjects with a computer player. Matching subjects with a computer player changes, however, two aspects: 1) social preferences are excluded – as subjects arguably cannot have preferences over payoffs of a computer – and 2) beliefs. Beliefs are changed as subjects might anticipate the computer player to be more rational or, alternatively, subjects might believe the computer to be more random in its decisions. To exclude the second aspect and to ensure that subjects' choices are driven only by social preferences and not beliefs, they were informed that computer players were imitating the behavior of other subjects. This means that the actions of the computer players were random draws from the set of human players' actions. This way, only social preferences should be impacted. However, a major downside of this controlled-belief manipulation is that the manipulation is very weak, as the spiteful preferences of the opponent are kept constant between treatments. The factor Opponent was realized via between-subjects design, i.e., subjects either interacted with a human player or a computer player imitating a human player.

Figure B.5 depicts the aggregate results. We see that both litigation expenditures and settlement requests are higher if participants compete against a fellow human compared to a computer. However, these differences do not rise to the required significance levels, and consequently, we find no statistically significant differences in the litigation expenditures nor in the settlement requests. On average, subjects invest 5.29 tokens in litigation against fellow humans, compared to 5.22 tokens in litigation against computers. The difference is statistically not significantly different from zero ($t(1633)=$ -0.7, $p\geq$ 0.05). Concerning the settlement requests, we find that subjects request on average 5.15 tokens in case the litigation is born out against a human, compared to 5.01 tokens in the computer treatment. Again, we do not find a significant effect, using a t-test: $t(1632.4)=$ -1.4, $p\geq$ 0.05. Also, using a mixed-effect regression, reported in Table B.14, does not show any significant difference between the two treatments, even though the effect of the human-treatment is consistently positive.

Figure B.5: Average litigation expenditures and settlement requests in each treatment



Note: The panel on the left depicts the litigation effort, while the panel on the right illustrates the settlement requests. The left two bars in each panel indicate the behavior under the English fee-shifting rule, while the two bars on the right indicate the response under the American fee-shifting rule in each panel. Red bars show the response if the opponent is a computer player, while blue bars show the response if the opponent is a human player. The error bars indicate the confidence intervals with the sample size on top and the mean below.

Table B.14: Mixed-effects regression of the litigation expenditures and settlement requests by opponent as a function of q

	Dependent Variable:							
	Litigation/ Settlement							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	5.22***	5.01***	4.80***	4.99***	3.86***	3.98***	3.66***	4.02***
	(0.07)	(0.07)	(0.08)	(0.07)	(0.08)	(0.07)	(0.09)	(0.08)
Human	0.07	0.14	0.13	0.15	0.13	0.11	0.15	0.10
	(0.11)	(0.10)	(0.11)	(0.10)	(0.12)	(0.10)	(0.13)	(0.11)
English			0.84***	0.06			0.40***	−0.09
			(0.06)	(0.04)			(0.09)	(0.07)
Human:English			−0.11	−0.03			−0.04	0.01
			(0.08)	(0.06)			(0.13)	(0.09)
NumQ					2.72***	2.07***	2.29***	1.92***
					(0.08)	(0.05)	(0.10)	(0.08)
Human:NumQ					−0.12	0.06	−0.05	0.10
					(0.11)	(0.08)	(0.15)	(0.11)
NumQ:English							0.88***	0.30***
							(0.15)	(0.11)
Human:NumQ:English							−0.14	−0.09
							(0.21)	(0.15)
Litigation	✓	×	✓	×	✓	×	✓	×
Observations	16,350	16,350	16,350	16,350	16,350	16,350	16,350	16,350

Models (1), (3), and (5) estimate the litigation expenditures. Models (2), (4), and (6) estimate the settlement requests. Models (1) and (2) estimate the average effect of the manipulation on behavior. Models (3) and (4) estimate the average interaction effect of the manipulation and the merit of the case. Models (5) and (6) estimate the interaction effect of the manipulation, the merit of the case, and the fee-shifting rule. Standard errors are shown in parentheses. NumQ indicates the merit of the case while Eng denotes a dummy with value one for the English rule and zero for the American rule. Human denotes a dummy with value one if the opponent is a human player and zero if the opponent is a computer player. $^{+}p<0.1$; $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

We provide several explanations for why we do not find any significant differences between the human treatment and the computer treatment:

First, we conduct the experiment online. However, in online experiments, social preferences might not be as salient as in laboratory experiments. Participants have no way of meeting the other participants, nor do they feel very connected to them. Thus, social preferences might have played already a relatively small role compared to a real-world setting. This, in turn, might diminish the scope for the effects of spite between the treatments.

Second, we exclude social preferences altogether. That, however, also means we exclude not only spiteful preferences but also prosocial preferences and other social preferences like inequality aversion. These potentially counteracting preferences might cancel each other out and substantially undermine the overall effect of spite on the observed behavior. Thus, other social preferences might mask the effect of spite. Even though we find that the results have the right tendency, indicating that spite matters, the missing significance might be due to other social preferences, resulting in a weak manipulation.

Third, we keep the beliefs about the behavior of the opponent constant across the human and the computer treatment. While this design choice seems to be essential to make the experiment clean, it also substantially reduces the scope of the manipulation. An optimal treatment would exclude the spiteful preferences of all participants in one treatment and retain them in the other. We, however, only exclude the social preferences of the decision-makers, while we keep the social preferences of the opponents constant. However, subjects are expected to change their behavior 1) due to their spiteful preferences and 2) due to the best response to the spiteful preferences of the opponents. By keeping the behavior of the opponent constant, we factually exclude the second channel. Thus, the scope of the manipulation is substantially reduced as we can only observe behavioral responses due to the first channel.

Overall, there are multiple arguments why our manipulation might not have been working or might have only a very limited effect on behavior. Still, even though we find no significant differences between the treatments, the consistent tendency of the results provides further support for the impact of spite on litigation and settlement behavior.

B.6 Additional Figures

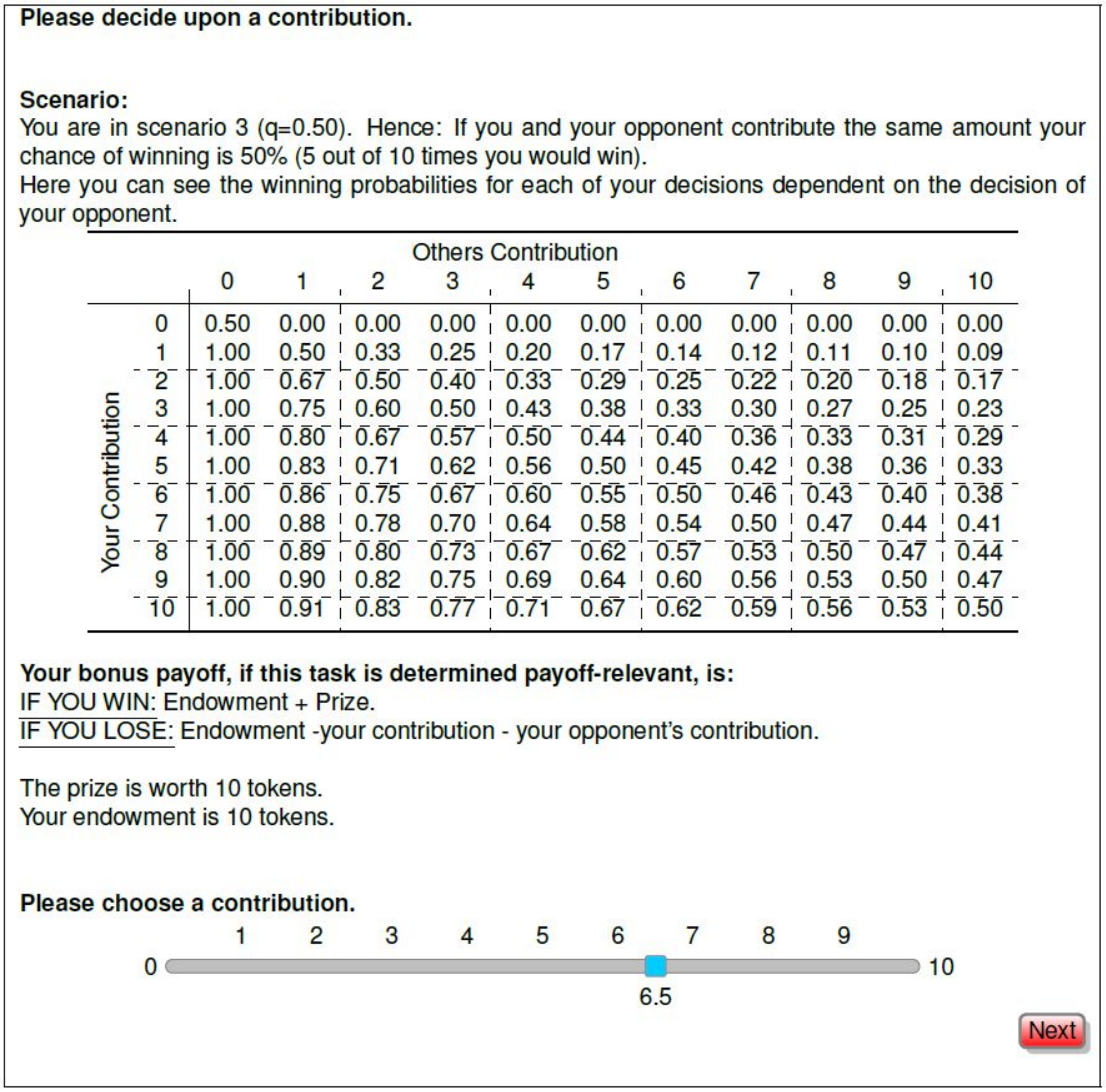
Figure B.6: Probability of winning litigation by fee-shifting rule as a function of q

Note: The panels on the top depict the winning probability of litigation by fee-shifting rule as a function of q, while the panels on the bottom show the aggregates. The panels in the first column show the winning probability of litigation of more (in blue) or less (in red) spiteful subjects when being matched with any of the other subjects, while the second and third columns illustrate the winning probability of litigation when being matched either only with less or more spiteful subjects. The error bars indicate the 95% confidence intervals.

Figure B.7: Probability of settlement by fee-shifting rule as a function of q

Note: The panels on the top depict the settlement probability by fee-shifting rule as a function of q, while the panels on the bottom show the aggregates. The panels in the first column show the settlement probability of more (in blue) or less (in red) spiteful subjects when being matched with any of the other subjects, while the second and third columns illustrate the settlement probability when being matched either only with less or more spiteful subjects. The error bars indicate the 95% confidence intervals.

Figure B.8: Interface for litigation expenditures under the English rule with $q = .5$



Please decide upon a contribution.

Scenario:
You are in scenario 3 (q=0.50). Hence: If you and your opponent contribute the same amount your chance of winning is 50% (5 out of 10 times you would win).
Here you can see the winning probabilities for each of your decisions dependent on the decision of your opponent.

Your Contribution \ Others Contribution	0	1	2	3	4	5	6	7	8	9	10
0	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.12	0.11	0.10	0.09
2	1.00	0.67	0.50	0.40	0.33	0.29	0.25	0.22	0.20	0.18	0.17
3	1.00	0.75	0.60	0.50	0.43	0.38	0.33	0.30	0.27	0.25	0.23
4	1.00	0.80	0.67	0.57	0.50	0.44	0.40	0.36	0.33	0.31	0.29
5	1.00	0.83	0.71	0.62	0.56	0.50	0.45	0.42	0.38	0.36	0.33
6	1.00	0.86	0.75	0.67	0.60	0.55	0.50	0.46	0.43	0.40	0.38
7	1.00	0.88	0.78	0.70	0.64	0.58	0.54	0.50	0.47	0.44	0.41
8	1.00	0.89	0.80	0.73	0.67	0.62	0.57	0.53	0.50	0.47	0.44
9	1.00	0.90	0.82	0.75	0.69	0.64	0.60	0.56	0.53	0.50	0.47
10	1.00	0.91	0.83	0.77	0.71	0.67	0.62	0.59	0.56	0.53	0.50

Your bonus payoff, if this task is determined payoff-relevant, is:
IF YOU WIN: Endowment + Prize.
IF YOU LOSE: Endowment -your contribution - your opponent's contribution.

The prize is worth 10 tokens.
Your endowment is 10 tokens.

Please choose a contribution.

Figure B.9: Interface for the settlement requests under the English rule with $q = .5$

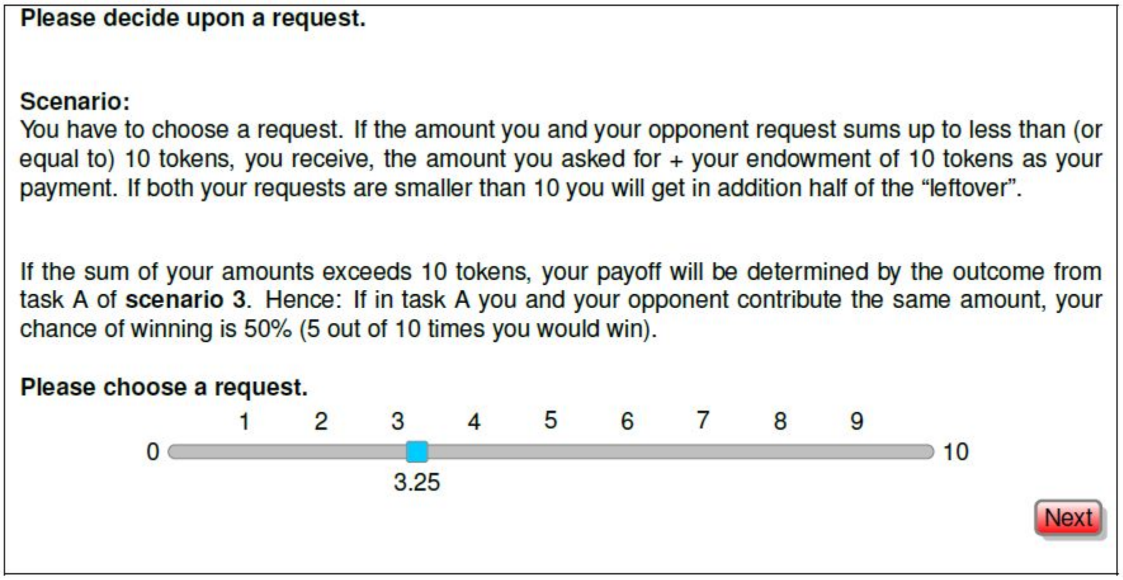
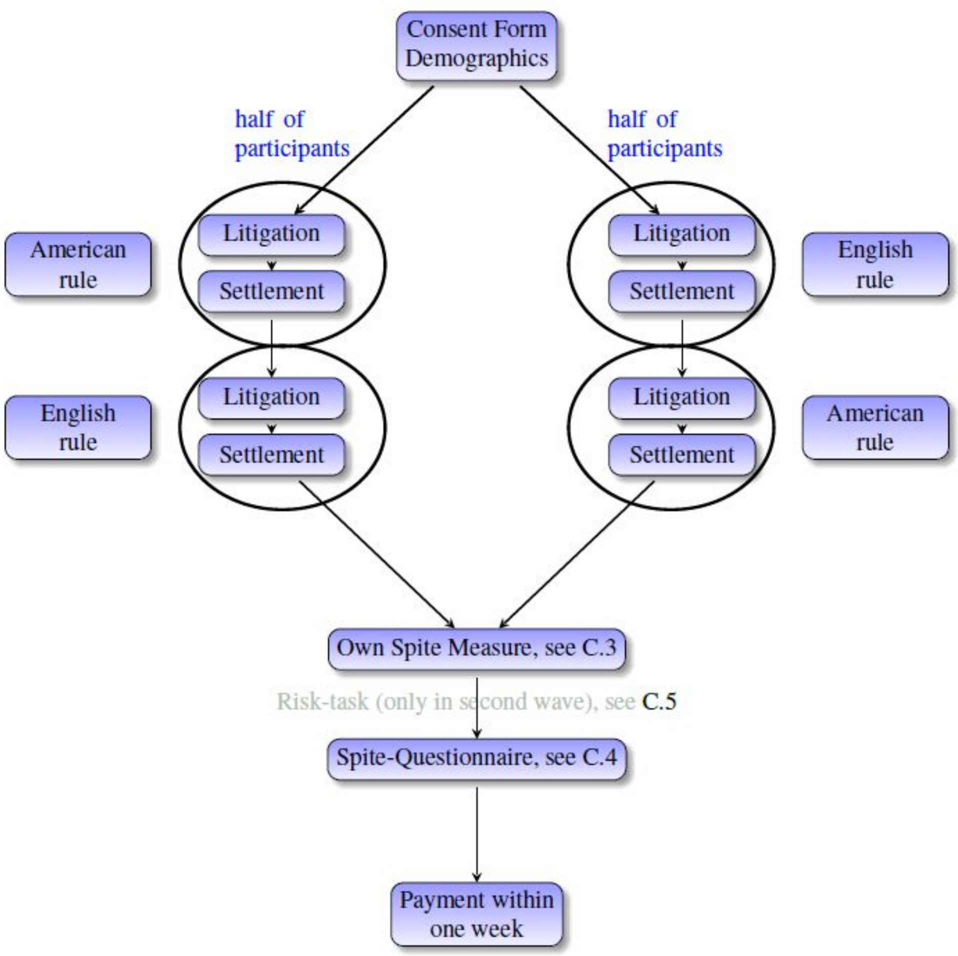


Figure B.10: Experimental procedure



B.7 Instructions and Control Questions

In the following, we show the instructions and control questions used in this experiment.

B.7.1 Instructions

The following depicts the instructions used in the experiment:

Welcome to this experiment in the economics of decision making.

If you follow these instructions carefully and make good decisions you will earn a considerable amount of money that will be paid to you within a few days to your MTurk account.

We ask that you pay close attention to the instructions.

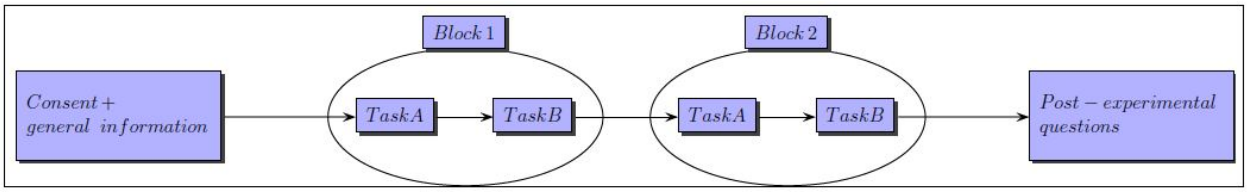
Note that during the experiment we will have several control question to see whether you read the instructions properly. If you read the instructions properly the control questions are very easy to answer. For every correctly answered control question, you will receive 5 cents in addition to your reward and your bonus payment from your decisions.

However, if you fail more than half of the control questions you will be excluded from all bonus payments and the experiment!

In the experiment today you will take decisions in two **blocks**.

Each **block** consists of two **tasks**. In both blocks you will need to make the same decisions; however, the blocks will differ in several aspects, which will be explained later in detail. The two tasks are either TASK A or TASK B. For each task, you will be instructed separately. Each task entails 5 decisions. Hence, overall you are going to make 5 (decisions per task) * 2 (tasks per block) * 2 (blocks) = 20 decisions.

The following graph illustrates the procedure of the experiment:



In the end, only one of the decisions (5 decisions in each task), from one of the tasks (two tasks in each block), from one of the blocks (two blocks) will be selected randomly.

Only this one selected decision will determine your payoff. Which one will be paid out was randomly determined when you agreed to take part in the study. However, you do not know which one will be payoff-relevant for YOU. Hence, you have to pay attention in each of the decisions as from your point of view any of the decision can be payoff-relevant for you.

Experimental Currency is used in the experiment. Your decisions and earnings will be recorded in tokens. Within a few days after the end of the experiment, you will be paid the bonus.

Tokens earned from the experiment will be converted to Dollars at a rate of:
1 token to 10 Dollar-cents ($0.10).

At the beginning of the experiment you are endowed with 10 tokens.
Any additional earning will be added to these tokens.
Any costs you encounter during your decisions will be deducted from the 10 tokens.

All tokens will be translated to dollars at the end of the experiment and paid as a bonus to you within a few days.

You have been assigned an opponent at the beginning of the experiment.
This opponent will stay your opponent for the duration of the whole experiment.
Importantly, the decisions of your opponent might influence your payoff.

[[in Computer treatment]]
Your opponent, however, is a computer player. This computer player will just copy the decisions of a real human player from a previous setting. Hence, the decisions of your opponent are implemented by a computer, but are copied from a human player. Your decisions can therefore **NOT** influence the payoff of your opponent, as the opponent is a computer player.

[[in Human treatment]]
All your decisions might also influence the payoff of your opponent, who is also a Mturker.

[[Instructions for the Litigation stage:]]

TASK A

In this task you are making a decision to win a prize worth 10 tokens. Your decision will influence your probability of winning this prize and hence your bonus payment.

Probability
For that purpose, you decide upon a contribution.
The higher your contribution the higher your chance of winning the prize. The higher the contribution of your opponent the lower your chance of winning the prize.

In addition: your chance of winning the prize does additionally depend on the scenario. The scenario describes your probability of winning the prize if both you and your opponent contribute the same amount.

Specifically, your chance of receiving the prize is given by your contribution divided by the sum of your contribution and your opponent's contribution as well as the scenario (q):

Chance of receiving the prize =

$$\frac{q \cdot (\text{your contribution})}{q \cdot (\text{your contribution}) + (1-q) \cdot (\text{your opponent's contribution})} \tag{B.14}$$

Where q represents the scenario and is a number between 0 and 1. The scenario describes your probability of winning the reward if both you and your assigned partner contribute the

same amount. Hence, it indicates whether the odds are in your favor.
Put differently: the scenario represents how much your contribution, relative to the contribution of your opponent, is weighted.

For example: if you and your opponent contribute the exact same amount and if the scenario is q = 0.5 then your chance of winning the reward is the same as your opponent's chance of winning. It also means, that your contribution has the same weight as the contribution of your opponent.

If however, you and your opponent contribute the exact same amount and the scenario is q = 0.9 then your chance of winning is 90 % and your opponent's chance of winning is 10 %, hence, the odds are in your favor. Put differently: your contribution is weighted 9 times more than the contribution of your opponent.

Another example: if you and your opponent contribute the exact same amount and if the scenario is q = 0.3 then your chance of winning is 30 % and your opponent's chance of winning is 70 %, hence, the odds are not in your favor. It also means, that one token of your contribution is weighted less than half (30/70) of one token of your opponent's contribution. Put again differently: to get the same odds of winning as your opponent, if your opponent contributes 3 tokens, you have to contribute 7 tokens.

Accompanying each scenario, you will see a simple table indicating your chance of winning in the respective scenario for possible contributions by you and your opponent.

The table will look like the following, which is an example table for scenario q = 0.90:
Note that you can choose any amount and for purpose of illustration we just pick integer (full numbers).

		Others contribution										
		0	1	2	3	4	5	6	7	8	9	10
Your Contribution	0	0.90	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.90	0.82	0.75	0.69	0.64	0.60	0.56	0.53	0.50	0.47
	2	1.00	0.95	0.90	0.86	0.82	0.78	0.75	0.72	0.69	0.67	0.64
	3	1.00	0.96	0.93	0.90	0.87	0.84	0.82	0.79	0.77	0.75	0.73
	4	1.00	0.97	0.95	0.92	0.90	0.88	0.86	0.84	0.82	0.80	0.78
	5	1.00	0.98	0.96	0.94	0.92	0.90	0.88	0.87	0.85	0.83	0.82
	6	1.00	0.98	0.96	0.95	0.93	0.92	0.90	0.89	0.87	0.86	0.84
	7	1.00	0.98	0.97	0.95	0.94	0.93	0.91	0.90	0.89	0.88	0.86
	8	1.00	0.99	0.97	0.96	0.95	0.94	0.92	0.91	0.90	0.89	0.88
	9	1.00	0.99	0.98	0.96	0.95	0.94	0.93	0.92	0.91	0.90	0.89
	10	1.00	0.99	0.98	0.97	0.96	0.95	0.94	0.93	0.92	0.91	0.90

The columns represent your opponent's contributions and the rows represents your possible contributions.
The numbers in this table represent your chances of winning, given yours and your opponent's contributions.
This table represents your winning probabilities in scenario q = 0.90.

For example: if both you and your opponent choose to contribute 2, your chance of winning is .90 (90 percent).
For example: if your contribution is 2 and your opponent's contribution is 6 your chance of winning is .75 (75 percent probability of winning the reward).

You will have to make a decision for each scenario.

There will be 5 scenarios.
The scenarios will be shown in random order.

Scenario1 (q = 0.10): If you and your opponent contribute the same amount your chance of winning is 10 % (you would win one out of 10 times)

Scenario2 (q = 0.30): If you and your opponent contribute the same amount your chance of winning is 30 % (you would win three out of 10 times)

Scenario3 ($q = 0.50$): If you and your opponent contribute the same amount your chance of winning is 50 % (you would win five out of 10 times)

Scenario4 ($q = 0.70$): If you and your opponent contribute the same amount your chance of winning is 70 % (you would win seven out of 10 times)

Scenario5 ($q = 0.90$): If you and your opponent contribute the same amount your chance of winning is 90 % (you would win nine out of 10 times)

Which scenario is relevant for your payoff was already determined before the experiment. However, you do not know which one will be payoff-relevant for YOU. Hence, you have to pay attention in each scenario as from your point of view any of the decisions can be payoff-relevant for you.

YOUR PAYOFF:

[[American rule]]
If you win you receive the prize and you will have to pay your contribution.
If you lose you will have to pay your contribution and you will NOT receive the prize.

[[English rule]]
If you win you receive the prize and you will not have to pay anything.
If you lose you will have to pay your contribution and you will have to pay the contribution of your opponent and you will NOT receive the prize.

<u>Hence your payoff is:</u>

[[American rule]]
IF YOU WIN: Endowment + prize -you contribution
IF YOU LOSE: Endowment - your contribution

[[English rule]]
IF YOU WIN: Endowment + prize
IF YOU LOSE: Endowment - your contribution - your opponent's contribution

Remember:
Your endowment at the beginning of the experiment was 10 tokens.
The prize is also worth 10 tokens.

Example:

Imagine, at the beginning of the experiment the first task was randomly selected to be payoff-relevant for you.

Imagine, of the first task the third scenario (q=.50) was randomly selected to be relevant for you.

Hence, your payoff is determined by your decision in this task, the decision of your opponent and a random draw. The third scenario is the scenario where your chance of winning the prize, if both you and your opponent contribute the same amount, is 50%.

The table explaining your winning probabilities given possible contributions of you and possible contributions of your opponent is given by:

		Others contribution										
		0	1	2	3	4	5	6	7	8	9	10
Your Contribution	0	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.12	0.11	0.10	0.09
	2	1.00	0.67	0.50	0.40	0.33	0.29	0.25	0.22	0.20	0.18	0.17
	3	1.00	0.75	0.60	0.50	0.43	0.38	0.33	0.30	0.27	0.25	0.23
	4	1.00	0.80	0.67	0.57	0.50	0.44	0.40	0.36	0.33	0.31	0.29
	5	1.00	0.83	0.71	0.62	0.56	0.50	0.45	0.42	0.38	0.36	0.33
	6	1.00	0.86	0.75	0.67	0.60	0.55	0.50	0.46	0.43	0.40	0.38
	7	1.00	0.88	0.78	0.70	0.64	0.58	0.54	0.50	0.47	0.44	0.41
	8	1.00	0.89	0.80	0.73	0.67	0.62	0.57	0.53	0.50	0.47	0.44
	9	1.00	0.90	0.82	0.75	0.69	0.64	0.60	0.56	0.53	0.50	0.47
	10	1.00	0.91	0.83	0.77	0.71	0.67	0.62	0.59	0.56	0.53	0.50

Imagine now that your opponent's contribution is one token and your contribution is one token. Hence, your chance of winning is 50%.

If you win (which happens in half of the cases) your payoff is:

[[American rule]]

Your Endowment + Prize -your contribution= 10+ 10 -1= 19 tokens

[[English rule]]

Your Endowment + Prize = 10+ 10 = 20 tokens

If you lose (which happens in half of the cases) your payoff is:

[[American rule]]	*[[English rule]]*
Your Endowment -your contribution = 10 -1 = 9 tokens	Your Endowment -your contribution - your opponent's contribution= 10 -1 -1 = 8 tokens

Imagine now that your opponent's contribution is three tokens and your contribution is one token. Hence, your chance of winning is 25%. Hence, in three out of four cases you would lose and in one out of four cases you would win.

<u>If you win (which happens in 1 out of 4 cases) your payoff is:</u>

[[American rule]]	*[[English rule]]*
Your Endowment + Prize -your contribution= 10+ 10 -1= 19 tokens	Your Endowment + Prize = 10+ 10 = 20 tokens

<u>If you lose (which happens in 3 out of 4 cases) your payoff is:</u>

[[American rule]]	*[[English rule]]*
Your Endowment -your contribution = 10 -1 = 9 tokens	Your Endowment -your contribution - your opponent's contribution= 10 -1 -3 = 6 tokens

Imagine now that your opponent's contribution is one token and your contribution is nine tokens. Hence, your chance of winning is 90%. Hence, in 9 out of 10 cases you would win and in 1 out of 10 cases you would lose.

<u>If you win (which happens in 9 out of 10 cases) your payoff is:</u>

[[American rule]]	*[[English rule]]*
Your Endowment + Prize -your contribution= 10+ 10 -9= 11 tokens	Your Endowment + Prize = 10+ 10 = 20 tokens

<u>If you lose (which happens in 1 out of 10 cases) your payoff is:</u>

[[American rule]]	*[[English rule]]*
Your Endowment -your contribution = 10 -9 = 1 token	Your Endowment -your contribution - your opponent's contribution= 10 -9 -1 = 0 tokens

[[Instructions for the Settlement stage:]]

TASK B

In the second task you will still be playing with the person assigned to you at the beginning of the experiment.

DECISION:
You and your opponent both can ask for a fraction of a prize. The prize is worth 10 tokens, just as in task A.

Payoff:
If the amount you and your opponent ask for sums up to less than (or equal to) 10 tokens, you receive, as payment, the amount you asked for. Hence, if the sum of both of your requests is smaller or equal to 10 tokens you will receive this requested amount as your payment plus your endowment. If both your requests are smaller than 10 you will get in addition half of the "leftover".

If the sum of your amounts exceeds 10 tokens, your payoff will be determined by the outcome from task A.

Hence, you will have to make again 5 decisions in the second task. Each decision is an amount you request from the 10 tokens. If both your requests are in sum less or equal to 10 this will be your payoff + half of the "leftover" + your endowment. If both your requests sum to more than 10 your payoff is determined by the result of task A.

EXAMPLES:
Imagine you request 3 tokens and your opponent requests 3 tokens. The sum is 6 and obviously smaller than 10. Hence, you will get as payoff your request (3 tokens) + half of the leftover (the leftover is 4 tokens) which is 2 + your endowment. Therefore, your total payoff equals to 15 tokens.

Imagine you request 3 tokens and your opponent requests 7 tokens. The sum is 10. Hence, you will get as payoff your request (3 tokens) + half of the leftover (the leftover is 0 tokens) which is 0 + your endowment. Therefore, your total payoff equals to 13 tokens.

Imagine you request 7 tokens and your opponent requests 7 tokens. The sum is 14. Hence, your payoff will be determined by the respective scenario from task one. Note that the range of total payoffs from the task A is 0 to 20 tokens.
For example, assume that the relevant scenario is q=0.10, assume also that you contributed in the first task 4 tokens and that your opponent contributed 4 tokens.

[[American rule]]
In case you win in task A (which would be the case in 1 of 10 cases given your contributions) your total payoff will be: your endowment + the prize - your contribution= 16 tokens.
In case you lose in task A your total payoff will be: your endowment - your contribution = 6 tokens.

[[English rule]]
In case you win in task A (which would be the case in 1 of 10 cases given your contributions) your total payoff will be: your endowment + the prize= 20 tokens.
In case you lose in task A your total payoff will be: your endowment - your contribution - your opponent's contribution= 2 tokens.

Before each decision, you will be told which scenario (q is either 0.1 or 0.3 or 0.5 or 0.7 or 0.9) from task one would be payoff-relevant if both your requests exceed 10 tokens.

B.7.2 Control Questions

The following control questions have been asked after the instructions of the litigation and the settlement decision.[4]

Litigation
Assume that task A (the task you just have been instructed to) has been randomly selected to be payoff-relevant for you.

Who is your opponent:

- (a) A fellow Mturker
- (b) A random computer
- (c) A computer imitating the choices of a previous participant
- (d) A fellow Mturker imitating the choices of a previous participant
- (e) Was not mentioned

[4]Note: in the second wave participatns were told that they would be able to proceed only if they answer all the questions correctly.

Assume that your contribution is 5 tokens and your opponent's contribution is 3 tokens and you win. What would be your total payoff?:

[[American rule:]]

- (a) 15 tokens
- (b) 10 tokens
- (c) 5 tokens
- (d) 25 tokens
- (e) 20 tokens

[[English rule:]]

- (a) 20 tokens
- (b) 10 tokens
- (c) 2 tokens
- (d) 25 tokens
- (e) 15 tokens

Assume that your contribution is 5 tokens and your opponent's contribution is 3 tokens and you lose. What would be your total payoff?:

[[American rule:]]

- (a) 15 tokens
- (b) 10 tokens
- (c) 5 tokens
- (d) 25 tokens
- (e) 20 tokens

[[English rule:]]

- (a) 20 tokens
- (b) 10 tokens
- (c) 2 tokens
- (d) 25 tokens
- (e) 15 tokens

Assume that your contribution is 1 tokens and your opponent's contribution is 3 tokens and you lose. What would be your total payoff?:

[[American rule:]]

- (a) 11 tokens
- (b) 9 tokens
- (c) 13 tokens
- (d) 19 tokens
- (e) 21 tokens

[[English rule:]]

- (a) 11 tokens
- (b) 6 tokens
- (c) 13 tokens
- (d) 20 tokens
- (e) 19 tokens

Assume that your contribution is 1 tokens and your opponent's contribution is 3 tokens and you win. What would be your total payoff?:

[[American rule:]]

- (a) 11 tokens
- (b) 9 tokens
- (c) 13 tokens
- (d) 19 tokens
- (e) 21 tokens

[[English rule:]]

- (a) 11 tokens
- (b) 6 tokens
- (c) 13 tokens
- (d) 20 tokens
- (e) 19 tokens

Imagine the payoff-relevant scenario for you is the third scenario (q=.50). Hence, your winning probabilities for receiving the prize are described by the following table:

		Others contribution										
		0	1	2	3	4	5	6	7	8	9	10
Your Contribution	0	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.12	0.11	0.10	0.09
	2	1.00	0.67	0.50	0.40	0.33	0.29	0.25	0.22	0.20	0.18	0.17
	3	1.00	0.75	0.60	0.50	0.43	0.38	0.33	0.30	0.27	0.25	0.23
	4	1.00	0.80	0.67	0.57	0.50	0.44	0.40	0.36	0.33	0.31	0.29
	5	1.00	0.83	0.71	0.62	0.56	0.50	0.45	0.42	0.38	0.36	0.33
	6	1.00	0.86	0.75	0.67	0.60	0.55	0.50	0.46	0.43	0.40	0.38
	7	1.00	0.88	0.78	0.70	0.64	0.58	0.54	0.50	0.47	0.44	0.41
	8	1.00	0.89	0.80	0.73	0.67	0.62	0.57	0.53	0.50	0.47	0.44
	9	1.00	0.90	0.82	0.75	0.69	0.64	0.60	0.56	0.53	0.50	0.47
	10	1.00	0.91	0.83	0.77	0.71	0.67	0.62	0.59	0.56	0.53	0.50

Suppose your opponent contributed 3 tokens. Suppose further that you contributed 7 tokens. What is your probability of winning the prize?

- (a) .50 (50% probability)
- (b) .70 (70% probability)
- (c) .84 (84% probability)

- (d) .88 (88% probability)
- (e) .90 (90% probability)

Suppose the same scenario is still payoff-relevant for you.
Suppose your opponent contributed 1 token. Suppose further that you contributed 1 token. What is your probability of winning the prize?

- (a) .50 (50% probability)
- (b) .70 (70% probability)
- (c) .84 (84% probability)
- (d) .88 (88% probability)
- (e) .90 (90% probability)

Settlement
Assume that task B (the task you just have been instructed to) has been randomly selected to be payoff-relevant for you.

Assume that your request is 5 tokens and your opponent's request is [[First question: 3]][[Second question: 5]][[Third question: 7]] tokens. Assume further that the relevant scenario is q=0.10. What would be your total payoff?

- (a) 16 tokens
- (b) 10 tokens
- (c) The payoff will be determined by the outcome from task A from scenario q=0.1
- (d) 15 tokens
- (e) 20 The payoff will be determined by the outcome from task A from scenario q=0.3

B.7.3 Own Spite Measure

In this task, you are still paired with your opponent from the previous tasks, whom we will refer to as the opponent. All of your choices will be confidential. After you take your decisions this task will not be repeated and there is no further interaction with your opponent.

You will be making a series of decisions about allocating resources between you and your opponent. For each of the following questions, please indicate the distribution you prefer most by selecting the button below the payoff allocations. You can only make one selection for each question. Your decisions will yield money for both yourself and your opponent.

Each point shown is worth 0.2 cents (100 points = 20 cents).

In the example below, a person has chosen to distribute the payoff so that he/she receives 50 points (=10 cents), while his opponent receives 40 points (=8 cents).

There are no right or wrong answers, this is all about personal preferences. After you have made your decision, select the resulting distribution of money by clicking on the button below your choice. As you can see, your choices will influence both the amount of money you receive as well as the amount of money your opponent receives.

At the end of the experiment, a computer program will randomly pick either you or your opponent as the payoff-relevant decision maker.

Only one of the following decisions will be payoff relevant. Which decision will be paid will be determined by a random process at the end of the experiment. Hence, you have to take all decisions seriously as any of those can be chosen by the random process with equal probability.

Your payment of this task will be added to your payment of the previous task.

Please indicate your choice for each of the following distributions.
Note: These decisions are payoff relevant and will influence your payment!
[[Participants had to make choices as shown in Table 2.1]]

B.7.4 Spite-Questionnaire

The questions of the questionnaire according to Marcus et al. (2014) included the following questions: 2

- I would be willing to take a punch if it meant that someone I did not like would receive two punches.
- I would be willing to pay more for some goods and services if other people I did not like had to pay even more.

- If I was one of the last students in a classroom taking an exam and I noticed that the instructor looked impatient, I would be sure to take my time finishing the exam just to irritate him or her.
- If my neighbor complained about the appearance of my front yard, I would be tempted to make it look worse just to annoy him or her.
- It might be worth risking my reputation in order to spread gossip about someone I did not like.
- If I am going to my car in a crowded parking lot and it appears that another driver wants my parking space, then I will make sure to take my time pulling out of the parking space.
- I hope that elected officials are successful in their efforts to improve my community even if I opposed their election. (reverse scored)
- If my neighbor complained that I was playing my music too loud, then I might turn up the music even louder just to irritate him or her, even if meant I could get fined.
- I would be happy receiving extra credit in a class even if other students received more points than me. (reverse scored)
- Part of me enjoys seeing the people I do not like fail even if their failure hurts me in some way.
- If I am checking out at a store and I feel like the person in line behind me is rushing me, then I will sometimes slow down and take extra time to pay.
- It is sometimes worth a little suffering on my part to see others receive the punishment they deserve.
- I would take on extra work at my job if it meant that one of my co-workers who I did not like would also have to do extra work.
- If I had the opportunity, then I would gladly pay a small sum of money to see a classmate who I do not like fail his or her final exam.
- There have been times when I was willing to suffer some small harm so that I could punish someone else who deserved it.
- I would rather no one get extra credit in a class if it meant that others would receive more credit than me.

- If I opposed the election of an official, then I would be glad to see him or her fail even if their failure hurt my community.

B.7.5 Risk Task

Here is a second short mini-experiment!
Another opportunity to earn money...

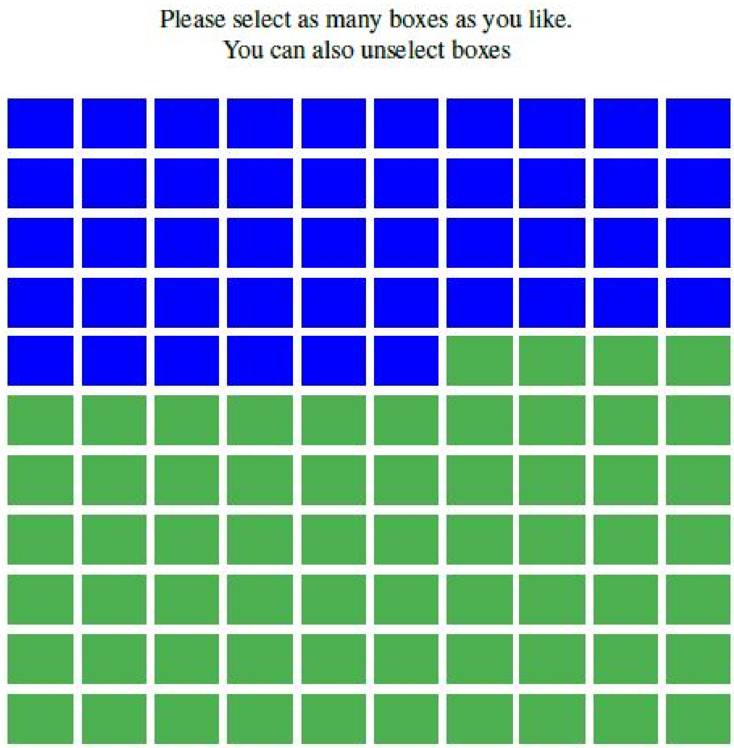
On this screen you will see a field composed of 100 boxes. Behind one of these boxes a bomb is hidden; the remaining 99 boxes are empty. You do not know where the bomb is. You only know that it can be in any place with equal probability.

Your task is to choose how many boxes to select. The position of the bomb will only be revealed after you made all your choices.

If you happen to have selected the box in which the bomb is located you will earn zero. If the time bomb is located in a box that you did not select you will earn 1 cent for each box you have chosen.

Below you will be asked to indicate which boxes you would like to select. You confirm your choice by hitting the next button. The position of the bomb will be revealed on the subsequent screen.

Figure B.11: Interface of the bomb task



Chapter 3

Social Norm Perceptions in Third-Party Punishment

with Katarína Čellárová

3.1 Introduction

In third-party punishment, an unaffected individual punishes another person for an act of wrongdoing. It is seen as a tool for the enforcement of social norms (Carpenter and Matthews, 2012, 2009; Henrich et al., 2006; Fehr and Fischbacher, 2004) and can serve to sustain cooperation by deterring selfish behavior (Lergetporer et al., 2014; Carpenter and Matthews, 2012; Mathew and Boyd, 2011; Charness et al., 2008; Carpenter et al., 2004; Fehr and Fischbacher, 2004), to promote more egalitarian allocations (Martin et al., 2021), and more generally, to sustain different norms of behavior across societies (Kamei et al., 2023; Henrich et al., 2006). Such punishment comes at a personal cost to the punisher, which suggests that humans care about how others behave in a specific situation, even when they are not directly affected by it.

Whatever constitutes wrongdoing, however, is subjective and not always clear, and it may depend on the punisher's perceptions of the relevant social norms. First, punishers may have their own personal beliefs about what should be done in a specific situation – *personal norms of appropriateness*[1] – and punish those who deviate from it. In this way, they could enforce their own preferences about how to behave in a specific situation. Second, because humans are a part of society, they may base their punishment decisions not only on their own

[1]We follow Bicchieri (2016) to classify the beliefs that matter for norm compliance. For a more concise text, we refer to personal norms as one of the ways in which social norms are 'perceived'.

appropriateness views but also on what others deem appropriate. *Normative expectations*[2] are individuals' beliefs about what others think is appropriate, and can guide punishment decisions to enforce behavior that individuals perceive to be preferred by society. Third, punishers may also rely on their *empirical expectations*[2], their beliefs about what constitutes common behavior. Empirical expectations may inform individuals' punishment decisions, as how humans typically behave may result from what they think others (and themselves) think is appropriate (Tremewan and Vostroknutov, 2021).

It is often argued that the existence of third-party punishment is evidence by itself that humans care about the enforcement of social norms (Carpenter and Matthews, 2012, 2009; Henrich et al., 2006; Fehr and Fischbacher, 2004). However, to the best of our knowledge, no study specifically elicits personal norms, normative expectations, and empirical expectations together and addresses their roles as the underlying motives for third-party punishment. Furthermore, no other study explicitly identifies the causal impact of the three types of social norm perceptions on third-party punishment decisions. In this paper, we close this research gap and identify whether and to what extent personal norms of appropriateness, empirical expectations, and normative expectations trigger third-party punishment and study their relative importance.

Previous studies indicate that social norms and individuals' beliefs about those norms matter for third-party punishment. Carpenter and Matthews (2009) test a broad set of different average behavior specifications in public goods games and find that deviations from the average contribution of the session best explain larger third-party punishment decisions.[3] Carpenter and Matthews (2012) confirm this result and further observe that deviations from the punisher's beliefs about the expected contribution, as well as from the punisher's own contribution, are associated with larger punishment decisions. Hence, in these studies, empirical expectations seem to matter for third-party punishment decisions. Moreover, other studies indicate that normative expectations matter as well. House et al. (2020) find that injunctive norm nudges in the form of messages about 'what is wrong and bad behavior' increase children's third-party punishment decisions. Zong et al. (2021) find that punishers react to information about the sender' expectations in a trust game. Dimant and Gesche (2023) observe that injunctive and descriptive norm nudges increase third-party punishment decisions and further find that both nudges increase personal appropriateness ratings of the situation in a subsequent experiment. However, in the subsequent experiment, they do

[2]Cialdini et al. (1990) defines the injunctive norm as what people believe ought to be and the descriptive norm as what usually is (common behavior). We follow the classification of Bicchieri (2016) and it can be viewed as the individual's belief of the injunctive norm (normative expectation) and the belief of the descriptive norm (empirical expectation).

[3]The set of behavior specifications included the average contribution of the session, of the ingroup, or of the outgroup, own contributions, or the set of all possible (exogenously set) contributions.

not elicit punishment decisions. Finally, personal norms also seem to matter. Bašić and Verrina (2023) find that both normative expectations and personal norms about third-party punishment decisions are positively correlated with own third-party punishment decisions.

Other studies also investigate other types of third-party punishment norms. Kamei (2020); Fabbri and Carbonara (2017), and Lois and Wessa (2019) find that information and beliefs about others' punishment decisions influence subjects' own punishment decisions. Furthermore, Kamei (2018) finds that being observed by another punisher increases punishment, indicating that subjects care about conforming to a punishment norm. Literature on second-party punishment also identifies the importance of the descriptive and injunctive norms of cooperation for punishment (Li et al., 2021; Reuben and Riedl, 2013), as well as the punishment norm itself (Li et al., 2021).

In summary, the literature demonstrates that perceptions of social norms matter for third-party punishment decisions. Yet, unlike our paper, none of the above-mentioned studies explicitly elicits all three social norm perceptions – personal norms, empirical expectations, and normative expectations – with punishment decisions together. Therefore, they cannot clearly identify the channels for punishment decisions. In principle, all of the three norm perceptions inform each other and hence are correlated (Tremewan and Vostroknutov, 2021). At the same time, these three norm-related perceptions can differ due to heterogeneous preferences and asymmetries in the availability and processing of information about appropriate or common behavior. Therefore, it is important to consider all of these norm-related beliefs because otherwise the effect of one of them could be wrongly attributed to another. Furthermore, none of the papers provides evidence for a causal effect of social norm perceptions on punishment.

We run an online experiment that consists of two phases. In the first phase – the Experience Phase – punishers go through a modified dictator game in different roles. The dictator starts with an endowment of 100 CZK and decides to transfer either 0, 10, 40, or 50 CZK to the receiver.[4] In the second phase – the Punishment Phase – subjects choose whether and how much to punish *another* dictator for their behavior in the same type of game via the strategy method. The to-be-punished dictator does not interact with the punishers in any other way. We measure personal norms, empirical expectations, normative expectations, and emotions before the Experience Phase and after the Punishment Phase.[5]

To create more pronounced heterogeneities in social norm perceptions, and to study their

[4]We omit the choices of 20 and 30 to force more extreme transfers and hence to have a higher potential of shifting norm perceptions. 100 CZK were approximately 4 EUR, which corresponded to a student wage of 50 minutes.

[5]We control for negative emotions, as they are an important driver of third-party punishment (Jordan et al., 2016a; Carpenter and Matthews, 2012; Nelissen and Zeelenberg, 2009).

causal effects, we employ four treatments with an exogenous variation of the Experience Phase. Participants are randomly assigned to the role of the dictator (*Dictator treatment*), receiver (*Receiver treatment*), observer (*Observer treatment*), or to the *Baseline* treatment.[6] In the Dictator and Receiver treatments, participants played one round of the same dictator game before the Punishment Phase. In the Observer treatment, participants observed a transfer from a dictator from an earlier session, and in the Baseline treatment, the Experience Phase was omitted.

The treatments have the potential to change social norm perceptions in the following way. First of all, the mere assignment to the roles of dictator and receiver may make subjects shift their perceptions of appropriate and common behavior in a motivated way. While dictators could tell themselves that lower transfers are more appropriate and common to justify their own low transfers, receivers may tell themselves the exact opposite. Second, subjects in the Receiver and Observer treatment receive an exogenous signal about typical behavior, which could make them update their norm perceptions. As receivers could also feel stronger emotions with the associated transfer, we explicitly control for emotions.

We find that the treatment assignment leads to substantive variations in how subjects perceive social norms: all three norm perceptions are different between the treatments. Additionally, the within-subject differences between the three norm-related perceptions are also shifted by the treatments. This allows us to study their causal impact and identify their relative contributions to punishment decisions. As the treatment manipulation worked, we will now present our main findings.

Our main findings are the following: We find that an increase in all three norm perceptions leads to higher punishment decisions individually. However, when studying their joint correlations, we find that the correlation of normative expectations with punishment reverses to a significant negative correlation. The positive effect of personal norms and empirical expectations on punishment prevails. In other words, we find that subjects who believe higher transfers are more appropriate and those who think that dictators typically transfer more punish more. At the same time, if they believe others deem higher transfers more appropriate, they punish less. Thus, we provide consistent evidence for a positive impact of personal norms and empirical expectations on punishment. Conversely, higher normative expectations are associated with lower punishment, when controlling for either one of the other norm perceptions.

To explore the negative relationship between normative expectations and punishment further, we analyze punishers' normative expectations relative to their own personal norms.

[6] We acknowledge that the Dictator and Receiver treatments are not independent from each other. However, as the acquired experience in the Experience Phase substantially differs between the two roles, and the choices of the dictators are exogenous to the receivers, we nonetheless classify them as separate treatments.

We find that subjects who believe to hold higher moral standards than society, punish more, whereas subjects who believe others to hold higher moral standards punish less. One explanation for this behavior is that individuals may feel a greater responsibility for punitive action when they anticipate lower moral standards among others, whereas they may not feel the necessity to enforce lower own appropriateness standards if they believe society to already uphold higher moral standards.

As additional results, we find that the relative importance of the three norm-related perceptions depends on gender and the assigned role. The positive relationship between personal norms and punishment is stronger for males, whereas the positive relationship between empirical expectations and punishment is stronger for females. Moreover, receivers rely more on their empirical expectations compared to the rest of the sample. Dictators hold the lowest personal norms compared to the other treatments, which results in overall lower punishment levels.[7]

The contribution of this paper is two-fold. First, we contribute to the literature on social norms and third-party punishment (Bašić and Verrina, 2023; Dimant and Gesche, 2023; Zong et al., 2021; House et al., 2020; Lois and Wessa, 2019; Kamei, 2018; Fabbri and Carbonara, 2017; Carpenter and Matthews, 2012, 2009; Henrich et al., 2006; Fehr and Fischbacher, 2004). We find a causal influence of social norm perceptions on punishment and thus provide evidence that third-party punishment is indeed used for the enforcement of social norms. We show that third-party punishment is used to enforce one's own view of appropriateness and typical behavior, but not society's appropriateness views. Furthermore, the importance of personal norms or empirical expectations varies depending on gender or the specific exogenously assigned role. These results provide policy implications for those aiming to increase informal sanctioning mechanisms, such as third-party punishment. Policies should focus on shifting empirical expectations (and, if possible personal norms) to influence punishment most efficiently. In addition, policymakers should evaluate who the target population is, as the relevance of the social norm perceptions for punishment varies.

Second, we also contribute to the literature about social norms (e.g. Bicchieri et al., 2023a; Abeler et al., 2019; Danilov and Sliwka, 2017; Kessler and Leider, 2012; Andreoni and Bernheim, 2009; Bénabou and Tirole, 2006). Our finding that punishment is not driven by the urge to enforce societal normative views has important implications for our understanding of norm-driven economic behavior more broadly. Individuals might overall care less about societal appropriateness views and more about their own appropriateness views and typical behavior. Additionally, we show the importance of considering all three norm-related beliefs.

[7] This indicates that dictators could engage in motivated reasoning. In order to licence themselves to transfer less, they lower their beliefs of appropriateness in a self-serving way.

As they are correlated but also differ from each other, the effect of one of them may be wrongly attributed to another one.

The paper is structured as follows: In Section 3.2, we provide a theoretical discussion, in Section 3.3, we describe the experimental design, and Section 3.4 presents the results. Lastly, Section 3.5 provides a discussion and concludes.

3.2 Theoretical Considerations

In this section, we discuss how social norm perceptions relate to each other and under which circumstances they may diverge. Based on this analysis, we derive the potential influence of social norm perceptions on third-party punishment decisions.

As Tremewan and Vostroknutov (2021) argue, individuals form social norm perceptions based on the information available to them and based on their perceptions of the availability of information to others. In the case of the personal opinion of appropriateness, individuals also take into account their own preferences about the expected outcomes that are associated with specific actions. Consequently, personal norms can differ among individuals due to differences in the information they rely on, different information processing, or heterogeneous preferences. When forming normative expectations, i.e., beliefs about others' personal views of appropriateness, individuals might compare others to themselves. If they believe others to be exactly alike, normative expectations coincide with their own personal norm of appropriateness. However, if individuals believe there are differences – either in the availability of information, information processing, or in preferences – normative expectations can differ from one's own personal norm of appropriateness.[8] Lastly, individuals' behavior may not necessarily align with their own or societal perception of appropriate behavior. Merguei et al. (2022) find that when there are several norms in a situation, subjects opportunistically follow the norm that maximizes their own payoffs – a phenomenon termed moral opportunism. Additionally, Bicchieri et al. (2023b) show that individuals motivatedly distort their social norm perceptions in a self-serving manner. They find that subjects update their empirical – but not normative – expectations about lying when presented with an upcoming opportunity to lie. Thus, individuals may choose to exploit these selfish opportunities or, crucially, expect others to do so. In addition, Kölle and Quercia (2021) show that when there is strategic uncertainty about others' behavior, normative and empirical expectations of participants differ substantially. Consequently, the perception of common behavior can differ considerably from normative views.

To sum up, the three norm-related beliefs can differ due to heterogeneities in preferences,

[8]See Bašić and Verrina (2023) for specific examples, where personal norms and normative expectations may differ.

due to differences in the availability and processing of information, and due to the anticipation of moral opportunism. Given that they differ, the question arises which among them mostly motivates individuals' third-party punishment decisions.

First, let's consider the enforcement of own personal norms. Punishing those who deviate from this view could be motivated by the desire to change future behavior. One goal could be to implement an outcome that one believes is better for everyone – or at least for individuals with the same type as themselves. Another goal could be to change future outcomes where one is directly involved. The reliance on personal norms goes in line with (Bašić and Verrina, 2023), who demonstrate the importance of personal norms for economic decision-making, including third-party punishment.[9]

Second, the desire to enforce an outcome that society deems appropriate may motivate punishment for the following reasons. For instance, individuals may see it as a moral obligation to serve society and punish those that deviate from the normative views that society imposes. Another possibility is that individuals may not have very strong own appropriateness views and thus generally rely more on societal appropriate views. If normative expectations and own personal norms diverge, punishers face a trade-off between enforcing an outcome that they deem appropriate and an outcome that, in their views, society deems appropriate. Whether the one or the other dominates punishment decisions may then depend on individual factors. Bašić and Verrina (2023) find overall stronger correlations of personal norms with economic behavior compared to normative expectations. This indicates that subjects might rely more on their personal norms than normative expectations when deciding on punishment.[10]

Lastly, previous literature emphasizes that empirical expectations are more important for economic behavior than normative expectations (Bicchieri et al., 2022; Kölle and Quercia, 2021; Chen et al., 2020b; Schmidt, 2019; Agerström et al., 2016; Bose et al., 2023; Bicchieri and Xiao, 2009) including for punishment decisions (Dimant and Gesche, 2023). Beliefs about common behavior may be used for punishment, if one wants to reinforce typical behavior by punishing those who behave atypically. In this case, subjects might decide to base punishment on empirical expectations instead of normative expectations because they want to justify and reinforce deviations from higher normative standards.[11] In addition, such a reliance on empirical expectations can also be used in a self-serving way to avoid the cost

[9]Unlike this study, (Bašić and Verrina, 2023) focus on appropriateness views about punishment decisions and not about behavior in the game itself.

[10]There might be cases when individuals engage in costly punitive behavior to enforce an outcome that they believe others would prefer, even if that goes against their own appropriateness views. Pluralistic ignorance – i.e., a difference between the perceived societal norm and all personal norms – is a prominent phenomenon that might be enforced in those cases (Andre et al., 2024; Bursztyn et al., 2020).

[11]This rationale also works when own personal norms are different from normative expectations.

of punishing others if one wants to behave opportunistically. This opportunity to decrease punishment may be especially exploited because empirical expectations are more prone to self-serving distortions compared to normative beliefs (Bicchieri et al., 2023b).

To identify the effect of each of these three norm-related beliefs, we aim to create heterogeneity in the appropriateness views and an additional mismatch between beliefs of common behavior and those appropriateness views. First, we use a (modified) dictator game, which is known to create substantial heterogeneities in behavior (Engel, 2011), which might be driven by different appropriateness views. As any allocation in the dictator game is Pareto-efficient, there is not one socially optimal solution. Hence, potentially heterogeneous fairness views and social preferences shape appropriateness views. Second, we employ four treatments to shift the three norm-related beliefs to a different extent. We do that in two ways: by assigning subjects to different roles in the game (additionally to their role of punishers) and by giving a noisy signal of common behavior in the form of one transfer decisions of a dictator.

Depending on the role that subjects get assigned to (receiver, dictator, or observer), they may motivatedly distort their beliefs in self-serving ways (e.g. Bicchieri et al., 2023b; Zimmermann, 2020; Epley and Gilovich, 2016). For instance, dictators may tell themselves that smaller transfers are appropriate, whereas receivers may believe that higher transfers are more appropriate. Additionally, participants may update their empirical expectations downwards or upwards around the reference point around the received signal of common behavior (Bicchieri et al., 2022; Hoeft et al., 2023; Gino et al., 2009; Keizer et al., 2008). At the same time, this signal may also inform normative expectations and change personal norms as they are related and inform each other.[12]

3.3 Experiment

We ran an online experiment[13] in March and April of 2021 with subjects of the Masaryk University Experimental Economics Laboratory (MUEEL). The experiment was programmed in z-Tree (Fischbacher, 2007), and we used z-Tree unleashed (Duch et al., 2020) to implement running sessions on the internet. The experiment received ethical approval from the GfeW.[14] Each session consisted of 14 participants and took an average of around 40 minutes, with average earnings of 118 CZK (approximately 4.79 EUR, which corresponded to a student wage of one hour of unqualified work). In total, 420 subjects participated in the experiment, of which 300 acted as punishers, and 120 as punishees. We analyze the punishment behavior

[12] See Tremewan and Vostroknutov (2021) for how social norm perceptions inform each other.

[13] Arechar et al. (2018b), for example, find that an interactive public goods game with and without punishment can be conducted very reliably online and produces similar behavioral patterns as in the laboratory.

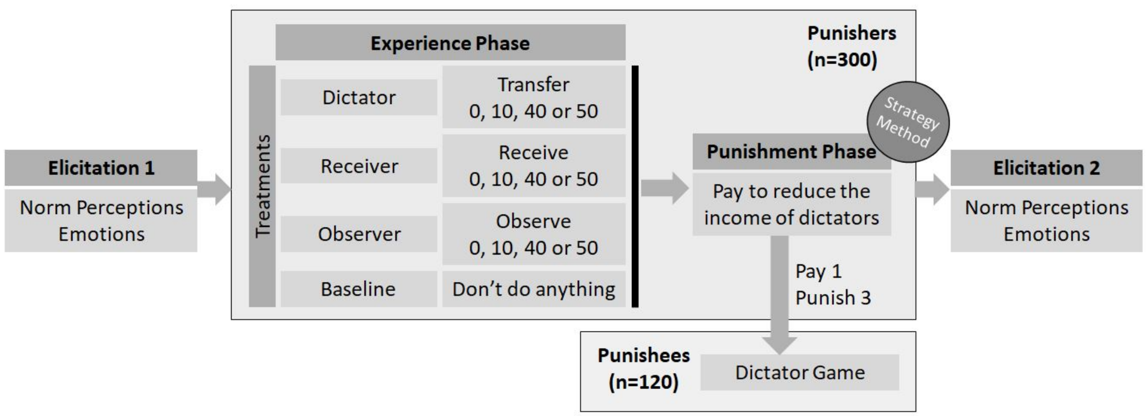
[14] German Association of Experimental Economics.

of 296 punishers.[15] Punishees played the dictator game and were subject to potential sanctions from the punishers, and they did not interact in any other way. Hence, we ensured impartial third-party punishment decisions and removed any indirect counter-punishment considerations.

3.3.1 Experimental Design

The experiment consisted of two sections, Section A and Section B, where Section A was payoff-relevant with an 80% probability and Section B with 20% probability. Section A was the main part of the experiment, whereas Section B served to measure distributive preferences and demographics. Figure 3.1 depicts Section A of the experiment. It consisted of two phases: the Experience Phase and the Punishment Phase. In addition, we measured social norm perceptions and emotions at the beginning and at the end of Section A.[16]

Figure 3.1: Experimental design Section A



In the Experience Phase, we manipulate what precedes the Punishment Phase. We employ four treatments with the goal to induce differences in the social norm perceptions. Participants were assigned to one of the following treatments: the *Dictator (N=80)* treatment, the *Receiver (N=80)* treatment, the *Observer (N=80)* treatment, and the *Baseline (N=60)* treatment.[17]

In the Dictator and Receiver treatment, which were conducted in the same sessions, subjects were randomly assigned to either the role of dictator or the role of receiver. Dictators

[15]We excluded four participants from the analysis because they remained inactive for several minutes and we had to forward them to the next pages.

[16]In principle, eliciting beliefs also at the beginning could overestimate the link between punishment and norm perceptions through an experimenter demand effect. However, d'Adda et al. (2016) do not find evidence that the order of norm elicitation affects behavior.

[17]We acknowledge that dictators and receivers were part of the same session and thus could also be classified as one Dictator/ Receiver treatment. As the Experience Phase is different between the two roles and the transfer is exogenous to the receiver, we call them as separate treatments.

decided how much to transfer to a randomly matched receiver. They were endowed with 100 CZK and could transfer either 0, 10, 40, or 50 CZK. In the Observer treatment, subjects observed the transfer of a randomly chosen dictator from the Dictator treatment, which was run in a previous experimental session.[18] In the Baseline treatment, the Experience Phase was simply omitted.[19] To avoid a systematic influence of income on punishment, we equalized payoffs by different show-up fees before the start of the Experience Phase. Receiver subjects were paid a show-up fee of 50 CZK. In the treatments Observer and Baseline, everyone received an individual show-up fee that consisted of 50 CZK plus a randomly chosen payoff from the set of payoffs that Receiver subjects had obtained in earlier sessions. Each payoff from this set was used exactly once for the Observer treatment. In the Baseline treatment, the distribution of payoffs was replicated.[20] We established the same wealth level before the Punishment Phase across all treatments except in the Dictator treatment, where subjects were wealthier by design.[21]

After this Experience Phase, the Punishment Phase followed. In the Punishment Phase, subjects could punish a dictator in the same version of the dictator game. This group of dictators (*punishees*, see Figure 3.1) was unrelated to the group of dictators in the Experience Phase and participated in the same sessions. We included punishees so that the punishment decisions of the punishers had real consequences. The Punishment Phase was the same for all treatments. All punishers could reduce the earnings of one punishee dictator. Punishers received an endowment of 50 CZK and could use that endowment to punish. We elicited the willingness to punish via the strategy method, where the punisher assigned deduction points to the punishee for every possible transfer.[22] We applied the typically used punishment ratio of 1:3, where the punisher pays one unit of her endowment to deduct the income from the punishee by three units (e.g. Fehr and Fischbacher, 2004).[23] If Section A was chosen to be payoff-relevant, the punishment decision was guaranteed to be implemented. We explain the exact matching procedure in the description of the punishees below. We made punishers

[18]Each dictator's decision was shown to one subject in the Observer treatment, who saw one specific decision. Dictators in the Dictator treatment did not know that their choices were shown to other players in later sessions.

[19]The Observer and Baseline treatments each took place in separate sessions.

[20]In the Baseline treatment we could not use the exact same number of payoffs because of a smaller number of subjects.

[21]In principle, a higher wealth could result in higher punishment. However, we find that wealth does not play an important role for punishment decisions.

[22]Jordan et al. (2016a) find that third-party punishment decisions are not influenced by the strategy method.

[23]It was possible to reduce the punishees' income by up to 150 CZK, which would result in a negative payment of the punishee. However, punishees played multiple rounds of the same dictator game. Therefore, negative payments in one round could be compensated by positive payments in another round, as well as by the elicitation stage and the show-up fee.

aware that they themselves would not be punished at any stage of the experiment.

In the whole experiment, we used a specific modified version of the dictator game. The dictator receives an endowment of 100 CZK, while the receiver starts with zero. The dictator can choose to transfer either 0, 10, 40, or 50 CZK to the receiver. This modified version has several advantages for studying the impact of social norm perceptions on punishment. By excluding intermediate choices like 20 or 30, we enforce more extreme transfers that have a larger potential to shift social norm perceptions of receivers and observers.[24] In addition, dictators have to choose a more extreme transfer in the Experience Phase. The need to justify a transfer of 0 or 10 over 40 or 50 could induce a more pronounced shift in norm perception through motivated reasoning. Another advantage of this modified version is that participants are less familiar with what describes common behavior and what ought to be done, which could lead to more heterogeneous social norm perceptions. Finally, as in the standard dictator game, there is no unique socially optimal allocation, and thus social norm perceptions may be more dispersed because subjects may deem different allocations as appropriate.

Before the Experience Phase and after the Punishment Phase, we elicited subjects' personal norms, normative expectations, and empirical expectations, following Bicchieri et al. (2022). First, we asked subjects what they believed *should* be transferred in the dictator game (personal norm). They could choose from the same set of transfers used throughout the whole experiment: 0, 10, 40, or 50 CZK. Second, we asked them what they thought was the *average response* to the first question by other participants of the same experimental session (normative expectation). Finally, we asked subjects *what they believed was the average choice* of the dictators in the ongoing (Dictator and Receiver treatment) or a previous (Observer treatment and Baseline) experimental session (empirical expectation). For both normative and empirical expectations, we used a continuous scale to capture small changes in individual norm perceptions.[25] The first norm elicitation took place after punishers knew their role in the Experience Phase. We incentivized the elicitation of normative and empirical expectations, paying an additional 15 CZK whenever participants were within a range of 6 CZK around the true average. In the second elicitation, which took place right after the punishment decision, subjects were shown their choices from the first elicitation. We asked them to consider whether their expectations had changed, thus, any reported change was intentional.

Lastly, we measured self-reported emotions following the elicitation of Bosman and

[24]In the pilot session, most of the initial norm perceptions were between 10 and 40 CZK. Thus, extreme transfers are further away from the initial norm perceptions and consequently have a larger potential to shift them.

[25]The initial position of the slider was 25, which constitutes half of the highest possible transfer.

Van Winden (2002) and Cubitt et al. (2011) and included both positive and negative emotions. Specifically, we asked participants about their current intensity of anger, gratitude, guilt, happiness, irritation, compassion, surprise, and envy. We measured the intensity of each emotion by self-reports on a 7-point Likert scale, from not feeling the emotion at all to feeling it very much. We elicited emotions both before the Experience Phase and after the Punishment Phase.[26] We included the elicitation of emotions as they have been shown to affect punishment (Jordan et al., 2016a; Carpenter and Matthews, 2012; Nelissen and Zeelenberg, 2009). At the same time, we control for any emotion that the Experience Phase induces, as to not wrongly attribute an emotion effect to a norm perception effect. This may be especially relevant for subjects in the Receiver treatment, who might experience strong emotions as the received transfer directly affects their payoff.

Next, we describe the procedure for punishees. Punishees started with the same elicitation of emotions and norms as the punishers, then played four rounds of the same version of the modified dictator game. In these four rounds, every punishee acted exactly twice as a dictator and twice as a receiver.[27] The exact rounds in which they acted in a specific role were randomly determined. When they were in the role of dictator, they were subject to punishment from the punishers, depending on the transfer that they chose.

In each round of the dictator game played by the punishees, we matched exactly one punisher with one round of a punishee dictator. Each session consisted of 14 participants, of which ten were punishers and four punishees. As each punishee acted exactly twice as a punishee dictator, the punishment decision of eight punishers was implemented. In this way, the punishment decision of every punisher was implemented if Section A was chosen for that punisher (since Section A was payoff-relevant with 80% probability).[28] Punishees knew that different punishers punished them in different rounds. However, we did not disclose this information to the punishers. We told punishers that punishees were in the same experimental session playing the same version of the dictator game and that punishers could reduce the income of a punishee dictator. Punishers were told about the implemented punishment decision and the choice of the punishee dictator only at the end of the experiment. Hence, the second norm elicitation of punishers remained unaffected by the transfer choice of the punishee dictator.

Finally, Section B served to measure distributional preferences. All punishers played the same dictator game as in Section A without the possibility of receiving punishment. All

[26] At the second elicitation, choices from the first elicitation were prefilled. We asked subjects to consider whether the intensity of their emotions had changed.

[27] We imposed the condition that each punishee would be twice in the role of dictator and twice in the role of receiver, to provide more equal payoffs for punishee participants.

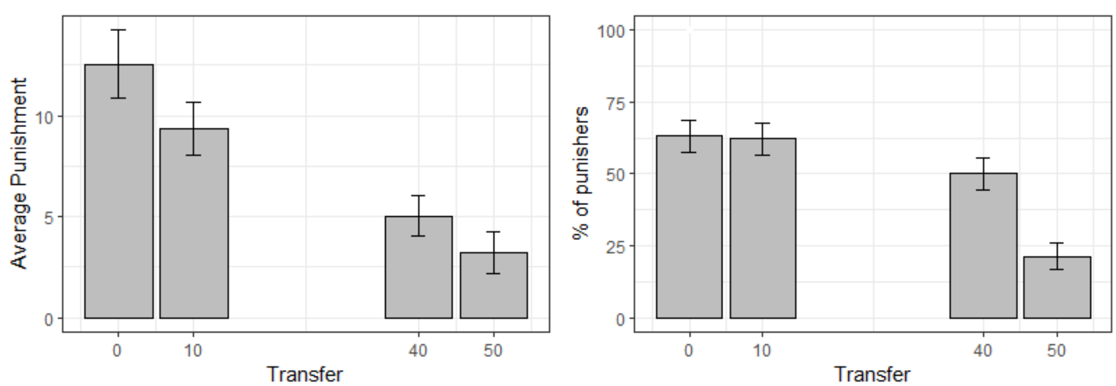
[28] Section A was chosen to be payoff-relevant for exactly eight punishers, while Section B was chosen to be payoff-relevant for exactly two punishers.

subjects made decisions as dictators. Afterwards, the computer decided randomly whether their role was a dictator or receiver and who they were matched with. Subjects then answered a questionnaire with demographic information, for which they received 30 CZK, in case Section B was payoff relevant (additional to a 50 CZK show-up fee). This part served as a control for differences in redistributive preferences, which we use as an additional control in a robustness check.

3.4 Results

In this section, we present the results of our experiment. Figure 3.2 depicts the overall punishment and propensity to punish all possible transfers of the dictator game elicited via the strategy method. Subjects mainly punish dictators, who transfer 0, 10, and 40 CZK. The amount of punishment decreases with the more equal splits, indicating that subjects want to enforce more equal allocations. In our analysis, we focus on prosocial punishment and treat the transfers of 0, 10, and also of 40 as such. Approximately 60% of all subjects punished a dictator for low transfers of 0 and 10, and around 50% for a transfer of 40. Thus, we consider the punishment of 40 as prosocial as well, as any deviation from the equal split of 50 may be perceived as selfish and less social behavior.[29]

Figure 3.2: Average punishment decisions and frequency of punishment



Note: The left figure shows the average amount of deduction points and 95% confidence intervals for every transfer that dictators could choose. The right figure depicts how many percent of subjects decided to punish a particular transfer at all and 95% binomial confidence intervals by the normal approximation method.

3.4.1 Social Norm Perceptions and Punishment

We use the treatment manipulation to increase heterogeneity among social norm perceptions by 1) assigning subjects to different roles and 2) receiving or observing different transfers.

[29] Around 20% of subjects also punished a dictator for a transfer of 50. This punishment can be considered as antisocial, which follows a different motivation.

Figure 3.3 shows the effects of the treatments on individuals' social norm perceptions and third-party punishment levels. It reveals substantive variations in the average norm perceptions between the treatments for the first and second elicitation of norms. In addition, the treatments cause substantive variations in the distribution of the norm perceptions (see Figure C.1 and Figure C.2 in Appendix C.1.1). Moreover, punishment decisions also vary across the treatments. What stands out from the figure is that when we compare the differences between the treatments, the punishment patterns closely resemble the patterns in social norm perceptions. This suggests a strong correlation which, in the following, we study more formally. First, we examine the patterns of social norm perceptions between and within subjects by pooling all treatments. Subsequently, we conduct a regression analysis. We focus on the second elicitation of norms, but all results can be replicated with the first elicitation of norms. Each individual is a statistically independent observation and hence our unit of analysis.

We find that average personal norms (mean = 29.80, sd = 19.4) are significantly ($p < 0.001$) higher than normative expectations (mean = 24.44, sd = 12.9), which are significantly ($p < 0.001$) higher than empirical expectations (mean = 19.57, sd = 12.8).[30] That means that, on average, individuals believe that what they themselves perceive as the appropriate transfer is higher than what others, on average, deem appropriate. The expectation of what is usually done is even lower.

The differences in those three norm-related beliefs are substantial also within subjects. Figure 3.4 depicts the individual-level differences between all pairwise combinations of the three social norm perceptions. While a difference of 0 is the most frequent for all three combinations, most of the density mass is not at 0.[31] Most subjects hold higher personal norms than normative expectations, higher normative expectations than empirical expectations, and also higher personal norms than empirical expectations.[32]

We find substantial differences between the three norm perceptions both aggregated and on a individual level. Thus, we are able to distinguish how each of the norm perceptions correlates with punishment. Now, we will examine whether punishment decisions are driven by what subjects believe should be done, what they believe others believe should be done, or what they think is usually done. We run regressions with all three norm perceptions and all subsets of combinations of the three as independent variables.[33] Table 3.1 shows

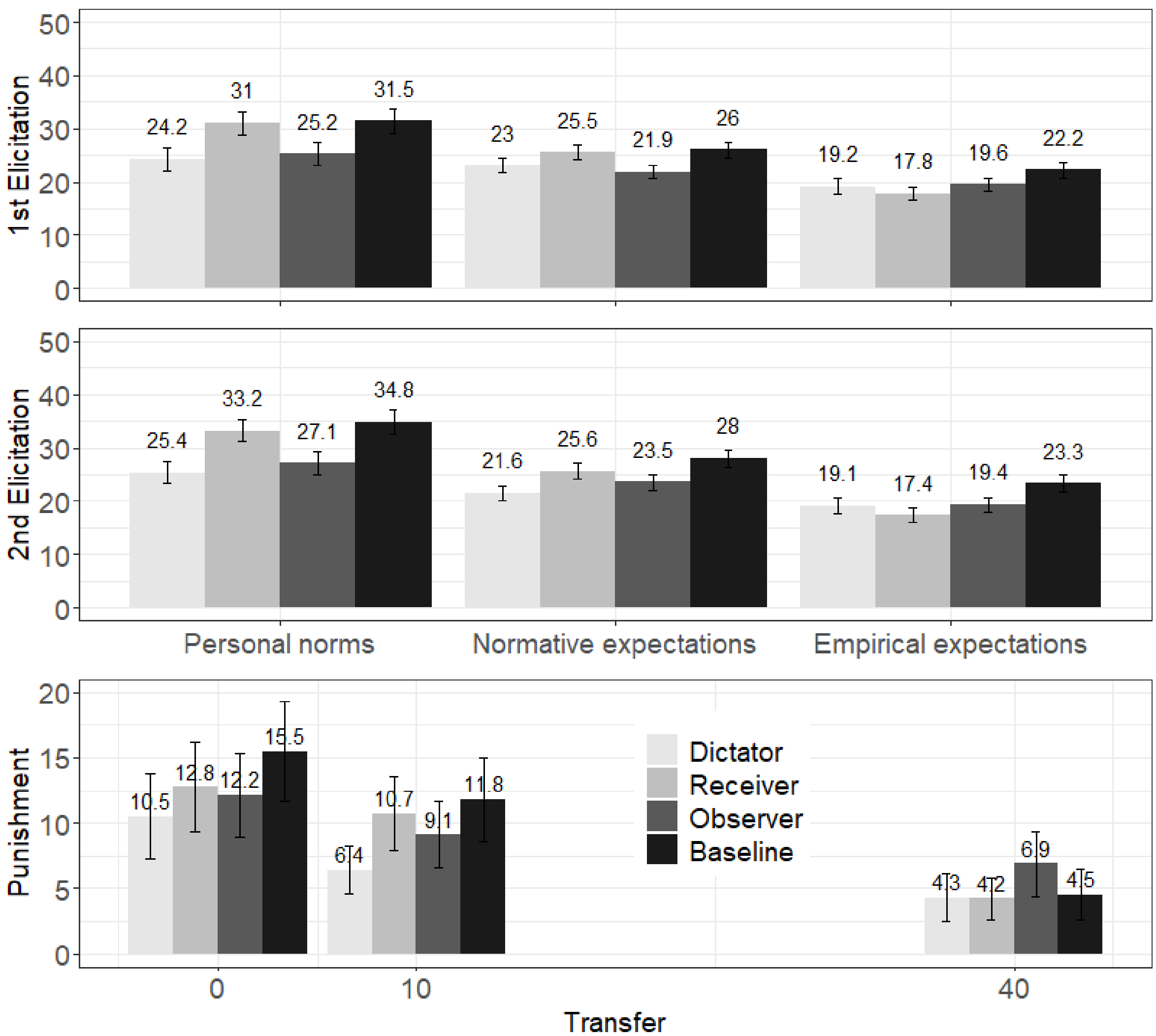
[30] All significance tests are non-parametric Wilcoxon signed-rank tests.

[31] The high peek at 0 for the gap between normative and empirical expectations is mostly driven by the Dictator treatment.

[32] The treatments also caused substantially different distributions of those individual differences (see Figure C.3 in Appendix C.1.1).

[33] In the robustness checks, we can replicate all results, when we define punishment as a function of how dictators deviate from the respective norms.

Figure 3.3: Social norm perceptions and punishment per treatment

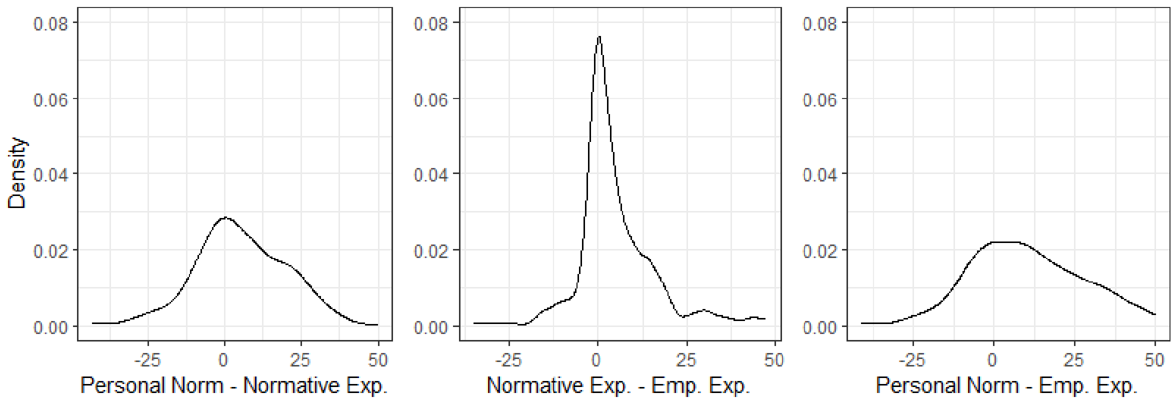


Note: The upper and middle panels show the three social norms perceptions conditional on the treatments of the first or second norm elicitation. The lower panel depicts punishment decisions conditional on the transfer of the to-be-punished dictator and the treatments. Error bars show 95% confidence intervals.

the results of these Tobit regressions (see Table C.1 of Appendix C.1 for the combinations of only two social norm perceptions).[34] In all models, we control for the dummy variables Transfer10 and Transfer40, which indicate the transfer of the to-be-punished dictator in the strategy method.[35] In addition, we control for negative emotions in all models to address potential endogeneity issues due to omitted variables, given that third-party punishment is known to correlate with negative emotions (Jordan et al., 2016a; Carpenter and Matthews,

[34]About 42% of all punishment decisions were 0, thus we use a Tobit model to account for such corner solutions.
[35]Thus, the baseline is for the punishment for a transfer of 0.

Figure 3.4: Within-subject differences in norm perceptions



Note: The figures show the Kernel densities of the within-subject differences in all pairwise combinations of the three norm perceptions (second elicitation).

2012; Nelissen and Zeelenberg, 2009).[36] The results do not substantially change and remain statistically significant if we exclude negative emotions.

In models (1), (2), and (3), we regress punishment decisions on each norm perception individually. In model (4), we include all three norm perceptions simultaneously, as they are highly correlated with each other (personal norms and empirical expectations: $\rho = 0.448$, personal norms and normative expectations: $\rho = 0.642$, and normative expectations and empirical expectations: $\rho = 0.626$, Spearman correlation, $p < 0.001$ all). This is important because the correlation between one of the norm perceptions and punishment without controlling for the other two would pick up the explanatory power of the others. By including all three simultaneously, we can study their relative importance.[37]

In the first three models, we find a positive and significant correlation between punishment and each of the norm perceptions individually. This changes when we include all three norm perceptions simultaneously. In model (4), we observe that the correlation of personal norms and empirical expectations with punishment decisions remains positive and significant. In contrast, the significant positive relationship between normative expectations and punishment decisions reverses to a significant negative relationship.[38] In other words, if subjects hold higher appropriateness views, or believe that higher transfers are more common,

[36] We declare a variable 'negative emotions' that consists of the average of anger, irritation, surprise, and envy (Cronbach's α is 0.69 ($CI_{95\%} = [0.66, 0.72]$) confirming the variable is internally consistent). We also include the differences in negative emotions of the second and first elicitation to capture emotional changes on the individual level caused by the Experience Phase.

[37] To deal with potential multicollinearity issues, we run regressions with pairwise subsets of all three norm perceptions in Appendix C.1.2 and run regressions with linear combinations of the variables in Section 3.4.2.

[38] In model (4), variance inflation factors are between 1.7 and 2.2 for all three norm perceptions. Hence, there is no indication that multicollinearity poses a serious problem for estimation.

Table 3.1: Tobit regression punishment on social norm perceptions

	Dependent Variable:			
	Punishment			
	(1)	(2)	(3)	(4)
Personal Norm	0.24***			0.23**
	(0.06)			(0.07)
Normative Expect.		0.19*		−0.25*
		(0.08)		(0.12)
Empirical Expect.			0.37***	0.36**
			(0.08)	(0.11)
Neg. Emotions	0.02	0.43	0.50	0.05
	(0.97)	(0.97)	(0.93)	(0.94)
Δ Neg. Emotions	2.11+	2.06	2.46+	2.45+
	(1.26)	(1.35)	(1.36)	(1.26)
TransferSM10	−4.23***	−4.19***	−4.24***	−4.26***
	(0.84)	(0.83)	(0.83)	(0.84)
TransferSM40	−11.52***	−11.45***	−11.50***	−11.56***
	(1.29)	(1.28)	(1.28)	(1.29)
Constant	0.97	2.47	−0.32	0.29
	(3.28)	(3.72)	(3.51)	(3.59)
Observations	888	888	888	888
# Clusters	296	296	296	296
Log Likelihood	−2,433.93	−2,450.44	−2,432.51	−2,419.70
Wald Test	96.45***	65.38***	100.35***	123.80***
	(df = 5)	(df = 5)	(df = 5)	(df = 7)

Note: SE clustered at individual level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

they consistently punish more independent of the other norm perceptions. This is different for normative expectations: holding personal norms and empirical expectations fixed, individuals who believe others hold higher appropriateness views punish less. This indicates that normative expectations carry a different motivation for punishment than personal norms of appropriateness and empirical expectations. To explore this motivation further, we study how the effect of normative expectations depends on their relative position with the two other norm-related beliefs in Section 3.4.2.

The observed dynamics remain similar when we regress combinations of only two of the three social norm perceptions (see Table C.1 in Appendix C.1.2). The coefficient of negative expectations, however, becomes close to zero and insignificant, when including only either personal norms and empirical expectations. The magnitude of the positive correlations between personal norms and empirical expectations with punishment is very similar across the models. This indicates that both matter for punishment decisions and explain a different part of the variation of punishment.[39]

We replicate our findings in all of the following robustness checks: First, we estimate the relationship between the first elicitation of norm perceptions and punishment decisions. Second, we estimate the relationship between both elicitations of norms and the propensity to punish. Third, we focus on deviations from the respective norms, i.e., the difference between the respective norm and the chosen transfer of the to-be-punished dictator. Fourth, we include the subject's choices of Section B of the experiment, where they themselves made a transfer decision as a dictator.[40] In all robustness checks, we can replicate the positive association of personal norms and empirical expectations with punishment and the negative association of normative expectations with punishment. Details can be found in Appendix C.1.3.

Lastly, we check whether our results are biased by the use of the strategy method. In particular, the relationship of empirical expectations with punishment could be biased because subjects might over-report punishment for transfers that they believe are unlikely to occur.[41] To check whether this is an issue, we run a robustness check (see Appendix C.1.4), where we declare a dummy that is one if the to-be-punished transfer is within a predefined neighborhood of an individual's empirical expectations. We find that empirical expectations are still significantly and positively related to punishment decisions after controlling for any of the neighborhood dummies. Hence, our results remain robust and the strategy method does not seem to affect our results substantially. Summarizing we find that:

Result 1. *Personal norms of appropriateness and empirical expectations are positively associated with punishment decisions.*

Result 2. *Normative expectations are negatively associated with punishment decisions.*

[39] Note that we elicit personal norms on a discrete scale (0, 10, 40, 50) and empirical expectations on a continuous scale (0-50). This leads to a measurement error for personal norms. With a positive correlation between personal norms and empirical expectations, this measurement error leads to an underestimation of the correlation between personal norms and punishment and an overestimation of the importance of empirical expectations.

[40] We do not find any significant association between this transfer and their punishment decisions.

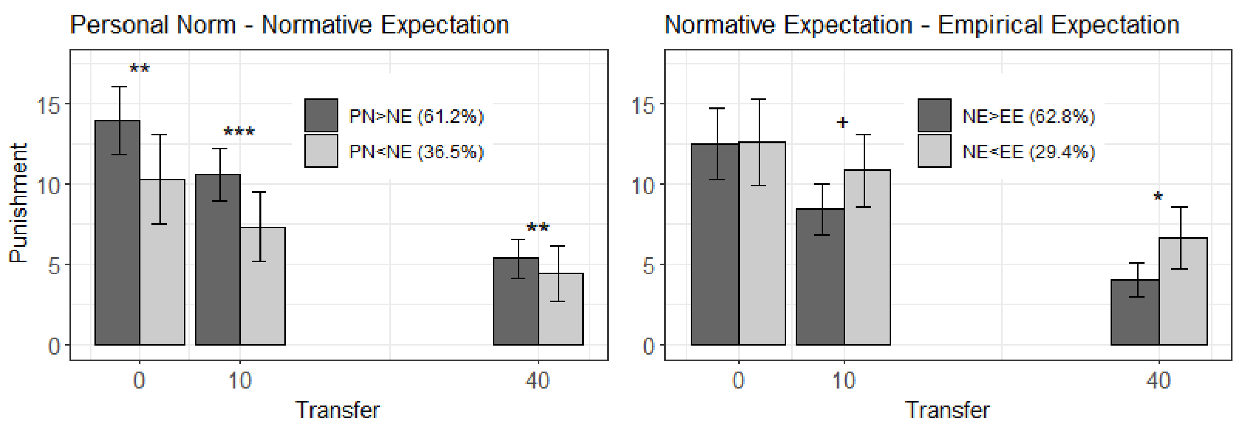
[41] With the strategy method, we elicit punishment decisions conditional on all possible transfers, but punishers only have to pay for the punishment of the actual transfer of the matched dictator. As a consequence, subjects with high empirical expectations might over-punish low transfers, as they do not expect to pay for this decision (and vice versa). This would lead to a stronger positive correlation between empirical expectations and third-party punishment.

3.4.2 Normative Expectations Gaps and Punishment

So far, we discovered that, surprisingly, normative expectations (beliefs about what others deem appropriate) are not positively associated with higher punishment decisions. This indicates that subjects do not punish because they want to enforce what they believe society deems appropriate. On the contrary, when they believe society holds higher moral standards, they even punish less. In this section, we explore the rationale behind this behavior and whether individuals consider normative expectations in relation to the two other norm-related perceptions. In particular, we analyze how the differences between normative expectations and the two other norm perceptions relate to third-party punishment decisions.

In a regression analysis (see Appendix C.2), we find a significant positive association between punishment and the difference between personal norms and normative expectations ($Gap_{PN-NE} = PN - NE$) and a significant negative association between punishment and the difference between normative and empirical expectations ($Gap_{NE-EE} = NE - EE$). Figure 3.5 illustrates those punishment differences conditional on the sign of these gaps, i.e., conditional on whether participants hold higher or lower normative expectations compared to either their personal norms (left graph) or empirical expectations (right graph).

Figure 3.5: Normative expectation gaps and punishment



Note: The left panel shows punishment decisions conditional on the transfer of the to-be-punished dictator for subjects with either higher (PN > NE) or lower (PN < NE) personal norms than normative expectations. The right panel shows these punishment decisions for subjects with either higher (NE > EE) or lower (NE < EE) normative than empirical expectations. The figure omits individuals with equal norm perceptions (PN=NE, 2.4%; NE=EE, 7.8%). All norms are based on the second elicitation. Error bars show 95% confidence intervals. Significance levels based on non-parametric Wilcoxon rank sum tests: + $p < 0.1$; *$p < 0.5$; ** $p < 0.01$; *** $p < 0.001$

We start with describing the positive relationship between punishment and the gap between personal norms and normative expectations. A positive gap indicates that individuals believe that the appropriate transfer according to themselves is higher than what society views as appropriate. As depicted in Figure 3.5, subjects with higher personal norms relative to normative expectations punish more than those with higher normative expectations than personal norms. When personal norms are higher than normative expectations, subjects may feel a greater responsibility to punish, as they think societal standards are lower and others are less likely to intervene. Conversely, if normative expectations exceed personal norms, subjects may anticipate others to punish more and consequently free-ride on their punishment decisions. Although subjects in the experiment were not aware of other punishers, experiences in the real world could still induce an instinct of the importance of taking actions, depending on the relation between one's own and others' appropriateness views. In other words, subjects may instinctively feel compelled to take punitive actions if they hold higher normative standards than others, or conversely, may refrain from punitive action if they believe others will uphold higher moral standards.

Second, we explore the negative relationship between punishment and the gap between normative expectations and empirical expectations. If normative expectations differ from empirical expectations, individuals think that others do not behave according to the social norm of appropriateness. Most of the participants (approximately 63%) hold higher normative expectations than empirical expectations.[42] They believe that the transfer, which is considered socially appropriate, is higher than the transfer that is actually sent. They suppose that even though others hold high standards of behavior, they actually do not act like that, and this belief of disconformity leads to lower punishment. For almost 30% of the participants, normative expectations were lower than empirical expectations[43]. These individuals expect others to give more than what is socially appropriate and punish more than the rest. This could be driven by the fact that those individuals, on average, hold a personal norm that is even higher than their empirical expectations. The exact reasons behind such a combination of beliefs and why it leads to higher punishment, however, can only be speculated upon.

To conclude, normative expectations affect third-party punishment indirectly through its relative position to the other two norm perceptions. We find a strong and stable positive association between the gap of personal norm (what one approves of) and normative expectations (what one believes others approve of) with punishment. The association between punishment and the difference between normative and empirical expectations (what society approves of vs. what one believes is typically done) is less pronounced and seems to matter only for punishment of higher transfers (see Figure 3.5). Therefore, we conclude that the gap

[42] The respective averages of norm perceptions (for those with norm. exp > emp.exp.) are: pers. norm: 31.2, norm. exp.: 27.1, emp. exp.: 17.2.

[43] The respective averages of norm perceptions (for those with norm. exp < emp.exp.) are: pers. norm: 28.3, norm. exp.: 19.6, emp. exp.: 24.2.

between personal norms and normative expectations matters for third-party punishment.

Result 3. *The gap between personal norms and normative expectations is associated positively with punishment decisions.*

3.4.3 Heterogeneity in Norm-driven Punishment

In addition, we find that the relative importance of the three norm perceptions for punishment decisions differs between subjects based on the role (treatment) they were assigned to and on gender. For instance, subjects in the role of receivers in the Experience Phase, seem to rely more on their empirical expectations when deciding on punishment compared to the rest of the sample. Subjects in the roles of dictators, observers, and subjects in the baseline hold qualitatively the same relative importance of the norms as shown in the main part (for more details, see Appendix C.4.1). Subjects in the role of dictators hold the lowest personal norms – possibly because of a motivated self-serving belief distortion – and punish significantly less than those in the Baseline treatment. Furthermore, we find substantial differences in the relative importance of the particular social norm perception for punishment between males and females. Males base their punishment decisions on their personal norms, whereas females rely on their empirical expectations (for more details, see Appendix C.4.2). Hence, it seems that males do not care about enforcing a social norm but rather what they personally deem appropriate, in contrast to females, who seem to care about the enforcement of typical behavior.

3.4.4 Causality

In this section, we explore whether social norm perceptions *causally* affect punishment, and tackle the challenges of reverse causality and omitted variable bias. In principle, the punishment decision itself could shape social norm perceptions and thus create the challenge of reverse causality. For instance, punishers may want to provide a reason for the way they punish and thus report their social norm perceptions in line with their punishment choices. Additionally, other individual characteristics may be correlated both with the norm-related beliefs and punishment and thus create the challenge of omitted-variable bias.

We will now show that reverse causality and omitted variable bias seem to not play a major role in our analysis. Specifically, we compare social norm perceptions between the first and second elicitation, as well as between punishers and punishees. Additionally, we compare the size of the correlation of social norm perceptions with punishment between the first and second elicitation. Finally, we run an IV analysis and an additional robustness check with individual-level controls.

First, we compare the first and second elicitations of norm perceptions in the Baseline treatment, where subjects engage only in punishment between the two elicitations. Hence, any changes in the norm perceptions can be attributed to the Punishment Phase. We do not find significant differences for empirical expectations (from 22.22 to 23.25), but we do find statistically significant ($p < 0.05$) increases in personal norms (from 31.50 to 34.83) and normative expectations (from 25.99 to 28.03) between the two elicitations. When we condition these changes on the levels of punishment, these differences look very similar. Therefore, the Punishment Phase itself seems to slightly influence normative views (personal norms and normative expectations), i.e., it increases how much subjects think should be sent and how much they believe others think should be sent. However, it does not affect subjects' beliefs about typical behavior, i.e., empirical expectations. In the following paragraphs, we will show that reverse causality does not play a major role, as the effect of social norm perceptions on third-party punishment outweighs any effect from punishment on personal norms and normative expectations.

For that, we compare the correlations of norm perceptions with punishment between the norm elicitation before the punishment opportunity (first elicitation) and the norm elicitation after the punishment opportunity (second elicitation). We find that the relationships are in the same direction, similar in size, and at the same significance level for all norm perceptions with punishment between the first and second elicitation (all models reported in Appendix C.1, Table C.2).[44]

Furthermore, we compare social norm perceptions of punishers and punishees to study whether the knowledge of the upcoming punishment opportunity changes social norm perceptions before the punishment opportunity. Punishees do not have this punishment opportunity, and thus, their norm perceptions are not influenced by this.[45] We find no significant differences between punishers and punishees norm perceptions, indicating that the knowledge about the upcoming punishment opportunity does not change norm perceptions (see Table C.8 in Appendix C.3). This fact, together with the same correlations of the first and second norm elicitations with punishment, demonstrates that reverse causality likely does not play a major role.

In addition, we employ an instrumental variable approach. Here, we exploit the changes in norm perceptions caused by the Experience Phase and the treatment manipulation. We use the treatment assignment to the roles of dictator, receiver, and observer, and the received or observed transfers as instruments for all three norm perceptions. We find that the

[44]The same holds true for the associations between the first and second elicitation with a punishment dummy (if the punisher decided to deduct points at all).

[45]Punishees act half of the time as dictators and the other half as receivers. Therefore, their norm perceptions should be the least biased of all subjects.

treatment manipulation (observing/ receiving a specific transfer or being assigned to the role of dictator) significantly shifts norm perceptions compared to the Baseline treatment (see Table C.9 in Appendix C.3.2). We find the most prominent changes in empirical expectations, followed by normative expectations and personal norms. For a more formal description and discussion of the IV approach, see Appendix C.3.2. Table 3.2 shows the results of the second stage of the Tobit IV for all three norm perceptions. In model (1), we use personal norms as an instrumented regressor, in model (2), normative expectations, and in model (3), we instrument for empirical expectations. In model (4), we instrument for all three norm perceptions simultaneously. In all models, we additionally instrument for negative emotions and the change in negative emotions. We replicate all results with models without negative emotions and the change in negative emotions (see Table C.10 in Appendix C.3.2).

Table 3.2: Tobit IV regression punishment on norm perceptions and negative emotions, second stage

	Dependent Variable:			
	Punishment			
	(1)	(2)	(3)	(4)
Personal Norms	0.62^{+} (0.36)			-0.14 (0.70)
Normative Exp.		0.75^{*} (0.37)		0.01 (0.87)
Empirical Exp.			0.80^{**} (0.30)	0.93 (0.59)
Negative Emotions	-3.93 (5.98)	-2.26 (5.76)	-3.53 (4.45)	-3.40 (5.20)
Δ Neg. Emotions	2.05 (9.20)	3.62 (8.59)	13.03^{+} (7.51)	14.70 (11.01)
Transfer10	-4.26^{***} (0.84)	-4.21^{***} (0.83)	-4.25^{***} (0.83)	-4.27^{***} (0.84)
Transfer40	-11.56^{***} (1.29)	-11.50^{***} (1.28)	-11.51^{***} (1.28)	-11.57^{***} (1.29)
Constant	0.21 (19.46)	-4.04 (18.94)	1.10 (12.56)	1.98 (16.36)
Observations	888	888	888	888
# Clusters	296	296	296	296
Log Likelihood	-8474.89	-8116.57	-8059.94	-14938.00
Wald χ^2 (df = 5)	88.33^{***}	90.29^{***}	90.04^{***}	90.76^{***}

Note: SE clustered at individual level and shown in parenthesis. Treatments conditional on transfer as instruments for empirical expectations, negative emotions, and Δ negative emotions. $+\ p < 0.1$; $^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$

Models (1), (2), and (3) show a statistically significant positive effect of each norm perception individually on punishment (weakly significant in the case of personal norms). As we show in Section 3.4.1, it is important to control for all three norm perceptions when estimating their relative effect. In model (4), however, we do not find a statistically significant impact of any of the three norm perceptions when instrumenting for all of them simultaneously. The reason for the failure in identifying their relative causal impact together is that all norm perceptions are shifted in the same direction in each of the treatments (see Regression Table C.9 in Appendix C.3.2). There is not enough induced variability between the three norm perceptions to disentangle the influence of each of them simultaneously in an instrumental variable regression.[46]

Finally, even though the IV analysis indicates that omitted-variable bias does not seem to play a role, we run an additional regression analysis with the following individual-level controls: age, gender, income, field of study, degree of understanding, and degree of concentration (see Appendix C.3.3). The relative correlations between the specific social norm perceptions with punishment remain robust to this additional model specification.

To conclude, given our analyses, we are confident that reverse causality and omitted-variable are rather unlikely to play a major role. Thus, we provide evidence that social norm perceptions affect third-party punishment *causally*.

3.5 Conclusion

In this paper, we show that perceptions of social norms matter for third-party punishment decisions. We explicitly measure three social norm perceptions about the behavior in a specific situation alongside punishment decisions, in contrast to the existing literature on social norms and punishment (e.g. Bašić and Verrina, 2023; Dimant and Gesche, 2023; Li et al., 2021; House et al., 2020; Reuben and Riedl, 2013; Carpenter and Matthews, 2012, 2009; Fehr and Fischbacher, 2004). We employ four treatments that manipulate subjects' perceptions of social norms and induce differences among them. This allows us to speak to their relative importance and to identify their causal effects on third-party punishment.

We find a consistent positive effect of personal beliefs about what should be done (personal norms) and beliefs about what is usually done (empirical expectations) on third-party punishment. This means that subjects who hold higher own moral standards, or believe that others typically behave more appropriately punish more. On the other hand, beliefs about what others believe should be done (normative expectations) are negatively correlated with punishment, when controlling for either personal norms or empirical expectations. This means that individuals punish less if they believe others to hold higher normative standards.

One explanation for this negative correlation could be that subjects anticipate interventions from others when they hold high normative standards, and thus would not have to intervene themselves. Conversely, when subjects believe others to hold lower normative standards, they anticipate fewer interventions and consequently feel compelled to enforce higher norms themselves. We provide evidence for this rationale, by finding that normative expectations matter in combination with own personal appropriateness views. Specifically, we find

[46] It proves challenging to isolate the specific causal effects of each social norm perception in this IV framework (controlling for all), as information provision leads to an update of all of them in the same direction. Future research should come up with instruments that move only the targeted norm perception while keeping the others constant.

that subjects whose normative expectations are higher than their personal beliefs punish less, whereas subjects whose personal norms are higher than their normative expectations punish more. Our argument aligns with the findings of Kamei et al. (2023), who observe that in the presence of other punishers, subjects tend to free-ride on others' punishment decisions.

Overall, our results show that the desire to enforce own beliefs of appropriateness and typical behavior motivates third-party punishment rather than perceived societal appropriateness views. Beliefs about societal appropriateness views seem only to matter in combination with personal norms, and could be used to determine the necessity of one's own punishment decisions. These findings extend and align well with previous literature on third-party punishment and social norms. Unlike existing studies, we provide a more complete picture of how all three norm-related beliefs motivate punishment choices.

For instance, our findings extend Carpenter and Matthews (2012) who find a positive correlation between the belief about average behavior (empirical expectations) with third-party punishment and Carpenter and Matthews (2009) who find that the session's averages best explain punishment compared to own group's averages, in- or out-group averages, own contributions, or the respective medians, minima, or maxima. On the other hand, our finding that third-party punishment is not driven by the belief of the injunctive norms – normative expectations – is in contrast with previous literature, for example with Bašić and Verrina (2023), who show that normative expectations about punishment choices are positively correlated with punishment, or House et al. (2020) and Dimant and Gesche (2023), who show that injunctive norm nudges increase punishment decisions. Instead, we even find a negative correlation when controlling for personal norms or empirical expectations. However, it goes in line with literature that finds that empirical expectations are more important for economic behavior than normative expectations (Bicchieri et al., 2022; Chen et al., 2020b; Schmidt, 2019; Agerström et al., 2016; Bose et al., 2023; Bicchieri and Xiao, 2009).

Our paper also adds to the literature about how normative and empirical information nudges affect third-party punishment. We find that our treatments shift social norm perceptions. For instance, we find that receiver and observers update their beliefs about common behavior depending on the dictator's transfer. Since appropriateness views are correlated with those beliefs, we also find an effect of the experienced transfers on normative expectations and personal norms of appropriateness. This fact can explain the results of House et al. (2020), Dimant and Gesche (2023), and Zong et al. (2021), who find that descriptive or injunctive norm nudges increase punishment decisions. Based on our results, the information about the descriptive or injunctive norm can change not only normative expectations but also empirical expectations and personal norms. Consequently, they impact third-party punishment decisions.

Our results also extend and possibly explain the results of Bašić and Verrina (2023); Kamei (2020); Lois and Wessa (2019); Fabbri and Carbonara (2017) who show that punishment decisions are correlated with beliefs and information about others' punishment decisions. Others' punishment decisions may also inform the injunctive and descriptive norm of the situation itself. Hence, punishers' beliefs about others' punishment decisions may correlate with their empirical expectations of the situation itself and, through them, correlate with punishment decisions.

As additional results, we find that the reliance on a specific norm-related perception depends on gender. Males punish primarily based on what they personally believe constitutes appropriate behavior. Females, on the other hand, primarily punish according to what they believe constitutes common behavior. The stronger reliance on empirical expectations for females contrasts with Croson et al. (2010) who find that males rely more on their empirical expectations compared to females when donating money. On the other hand, Fišar et al. (2016) do not find any gender differences in the relationship between empirical expectations and third-party punishment decisions in their study of bribing.

Furthermore, we also show that the importance of a specific norm perception and punishment depends on the mere assignment to a role. For example, being assigned to the role of receiver leads to a higher reliance on empirical expectations compared to the rest of the sample. Additionally, being assigned to the role of dictator leads to significantly lower social norm perceptions and, consequently, to lower punishment. This motivated shift of social norm perceptions is in accordance with the literature on motivated beliefs (e.g. Bicchieri et al., 2023b; Zimmermann, 2020; Epley and Gilovich, 2016).

We can draw policy recommendations from our results. Policies that are aimed at changing empirical expectations rather than normative expectations have a higher potential to change third-party punishment decisions. For instance, providing information about common behavior instead of what others deem appropriate may influence empirical expectations more than normative expectations and hence have a higher potential to shift punishment behavior. This would align with Dimant and Gesche (2023), who show that empirical information changes behavior more than normative information (although non-significantly). Additionally, as the reliance on either personal norms or empirical expectations depends on gender and the role that subjects are assigned to, specific information policies should be tailored to the needs of the specific audience in order to increase its effectiveness.

To conclude, we provide consistent evidence that social norm perceptions motivate third-party punishment. Individuals, who hold higher personal appropriateness views and believe that others behave more appropriately, punish more. On the other hand, subjects who believe others to hold higher normative views, punish less. In addition, we find that the initial positive correlation between normative expectations and punishment reverses when controlling for either of the other two norm perceptions. This has important consequences for the overall social norm literature. We show that it is important to consider all norm-related beliefs because otherwise, the effect of one of them might be wrongly attributed to another.

Acknowledgements for Chapter 3

We are grateful to Rostislav Staněk, Henrik Orzen, Wladislaw Mill, Marco Faillo, James Tremewan, Ondřej Krčál, Aleksandra Khokhlova, and Dam Thi Anh for their valuable comments. We appreciate comments from participants of the seminars of Ca'Foscari University of Venice, the University of Verona, and the CEEL University of Trento, the ZEW/Uni Mannheim Experimental Seminar, the CDSE Seminar in Mannheim, the International Online Conference RExCon21 on 'Social Preferences and Social Norms', the ESA GOACM 2021, and the Summer School of Behavioural Game Theory in Norwich. The financial support of the grants MUNI/A/0931/2019 and MUNI/IGA/1364/2020 is gratefully acknowledged. This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences and by the Faculty of Economics and Administration at Masaryk University.

Appendix C

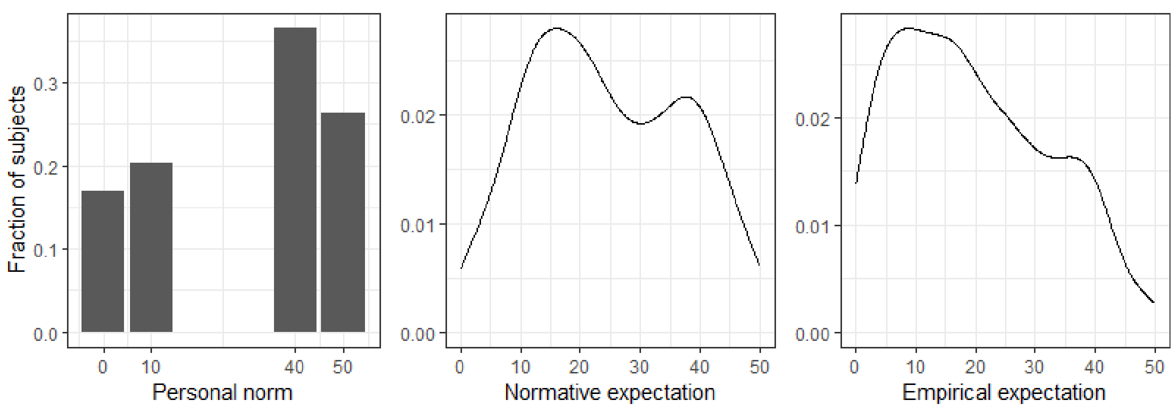
Appendix to Chapter 3

C.1 Social Norm Perceptions and Punishment

C.1.1 Distribution of Norm Perceptions

Figure C.1 shows the distribution of each norm perception at the second elicitation. It reveals that there is no clear consensus between subjects in all of the three norm perceptions.

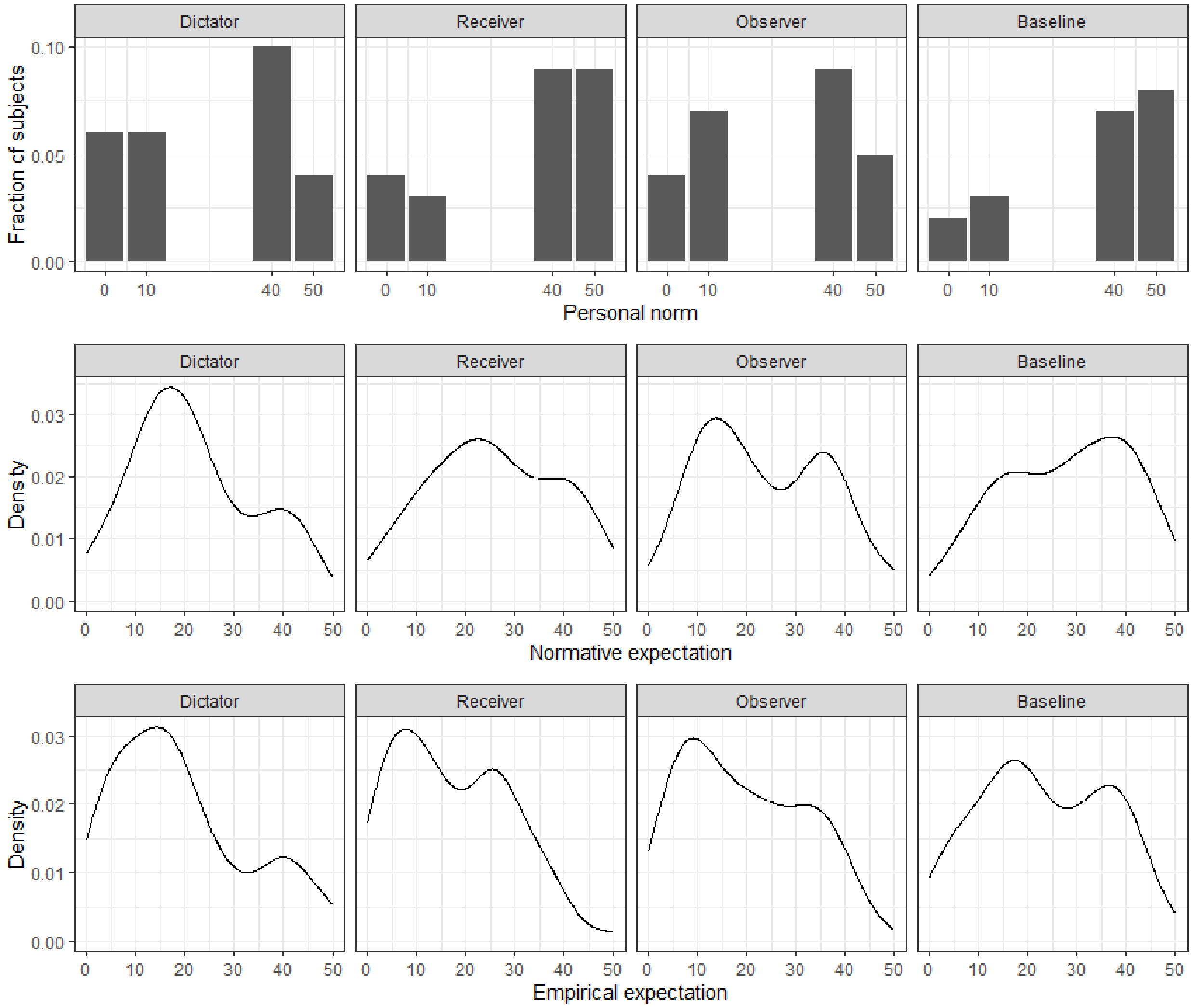
Figure C.1: Distributions of social norm perceptions in second elicitation



Note: The left plot shows the fractions of subjects who hold a particular personal norm. The plots in the middle and on the right show the distribution of normative and empirical expectations via Kernel densities.

Figure C.2 shows the distribution of each norm perception at the second elicitation conditional on the treatments. It shows that the treatment assignment leads to differences in the distributions of all three norm perceptions.

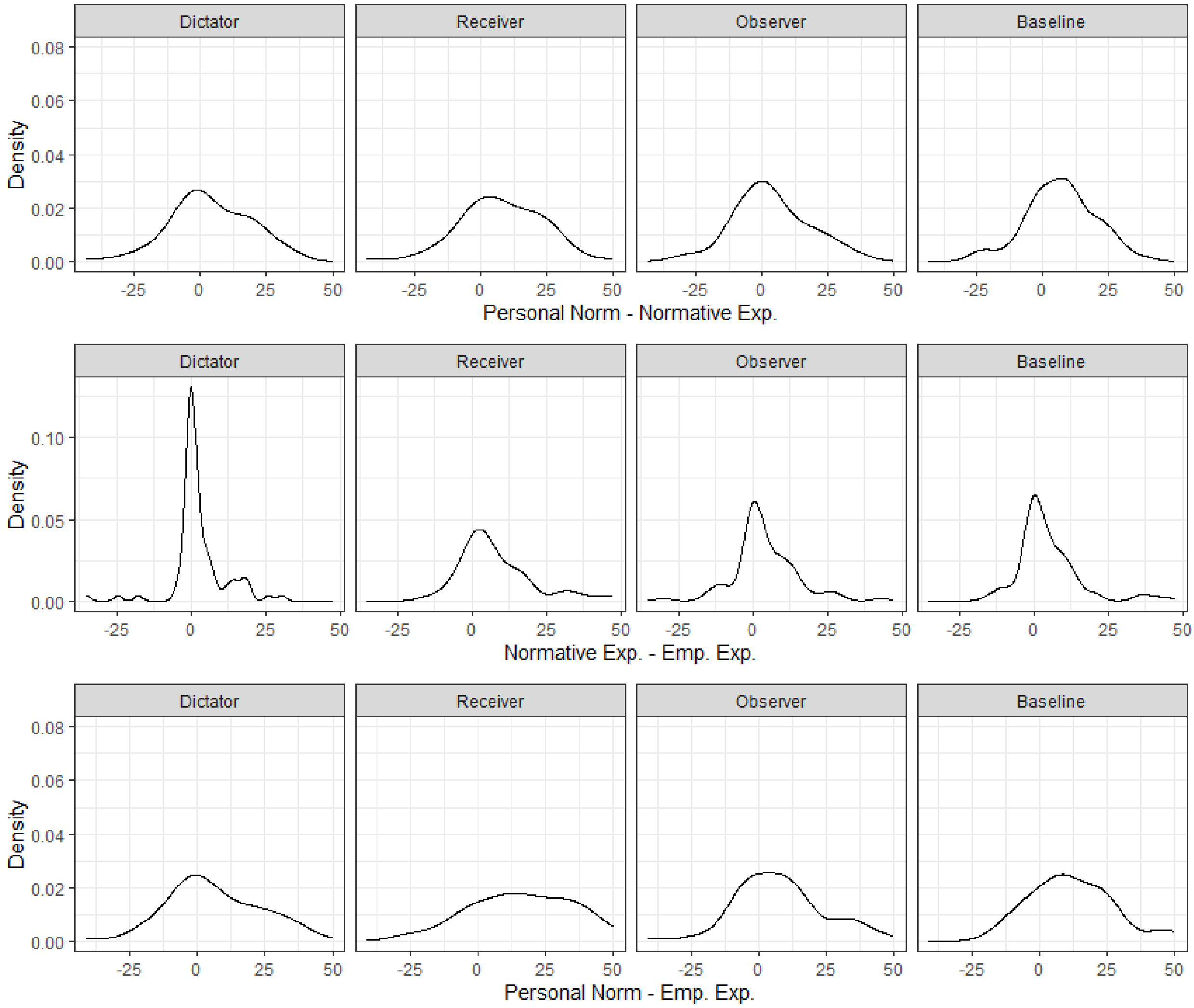
Figure C.2: Distributions of social norm perceptions in second elicitation per treatment



Note: The upper panels show the fractions of subjects who hold a particular personal norm conditional on the treatments. The middle and lower panels show the distribution of normative and empirical expectations via Kernel densities conditional on the treatments.

Figure C.3 shows the distributions of within-subject differences in the norm perceptions per treatment. It shows that the treatment assignment leads to differences in those distributions.

Figure C.3: Within-subject differences in norm perceptions per treatment



Note: The panels show the Kernel densities of the within-subject differences in the three norm perceptions (second elicitation) for all treatments separately.

C.1.2 Combinations of Norm Perceptions and Punishment

In this section, we regress punishment on combinations of social norm perceptions and punishment, where we take only two out of the three social norm perceptions. We replicate the results from the main text. Empirical expectations are significantly positively correlated with punishment, personal norms are significantly positively correlated, and normative expectations are (however non-significantly) negatively correlated with punishment.

Table C.1: Tobit regression punishment on social norm perception combinations

	Dependent Variable:		
	Punishment		
	(1)	(2)	(3)
Personal Norm	0.27^{***}	0.16^{*}	
	(0.07)	(0.07)	
Normative Expect.	-0.06		-0.07
	(0.10)		(0.11)
Empirical Expect.		0.26^{**}	0.41^{***}
		(0.10)	(0.11)
Constant	1.89	-2.18	0.56
	(3.63)	(3.50)	(3.65)
Transfer	✓	✓	✓
Neg. Emotions	✓	✓	✓
Δ Neg. Emotions	✓	✓	✓
Observations	888	888	888
# Clusters	296	296	296
Log Likelihood	$-2{,}433.54$	$-2{,}424.92$	$-2{,}432.03$
Wald Test (df = 6)	97.09^{***}	114.17^{***}	101.17^{***}
Note: SE clustered at individual level	$^{*}p<0.05$;$^{**}p<0.01$;$^{***}p<0.001$		

C.1.3 Robustness Norm Perceptions and Punishment

Next, we replicate the overall association between norm perceptions and punishment decisions. In Table C.2, in model (1), we regress punishment on the first elicitation of norm perceptions and in model (2) on the second elicitation. In models (3) and (4), we regress a punishment dummy on the first and second elicitation of norm perceptions. All models replicate the results from the main section: personal norms and empirical expectations are significantly positively associated with punishment decisions, while normative expectations are significantly negatively associated.

Table C.2: Tobit and Logit regression punishment on norm perceptions

	Dependent Variable:			
	Punishment		*Punishment Dummy*	
	Tobit		*Logistic*	
	(1)	(2)	(3)	(4)
	1st Elicitation	2nd Elicitation	1st Elicitation	2nd Elicitation
Personal Norm	0.18*	0.23**	0.02*	0.03**
	(0.08)	(0.07)	(0.08)	(0.07)
Normative Expect.	−0.36*	−0.25*	−0.04*	−0.04*
	(0.14)	(0.12)	(0.14)	(0.12)
Empirical Expect.	0.44***	0.36**	0.05***	0.04**
	(0.13)	(0.11)	(0.13)	(0.11)
Constant	2.07	0.29	−0.22	−0.24
	(3.76)	(3.59)	(3.76)	(3.59)
TransferSM	✓	✓	✓	✓
Neg. Emotions	✓	✓	✓	✓
Δ Neg. Emotions	✓	✓	✓	✓
Observations	888	888	888	888
# Clusters	296	296	296	296
Log Likelihood	−2,425.14	−2,419.70	−558.21	−545.02
Akaike Inf. Crit.			1,132.41	1,106.03
Wald Test (df = 7)	113.88***	123.80***		

Note: SE clustered at individual level $^{*}p < 0.05$;$^{**}p < 0.01$;$^{***}p < 0.001$

Next, we regress punishment on the deviations of the transfer to the norm perceptions. For this, we take the difference of each of the three norm perceptions to the to-be-punished transfer. Table C.3 reveals that the results remain robust to this model specification.

Finally, we include the subjects' own transfer decisions as a dictator in Section B of the experiment. Model (1) in Table C.4 reveals that the transfer in Section B is not significantly correlated with punishment decisions. Model (2) adds norm perceptions and shows that even after controlling for the transfers in Section B, the results from the main text prevail: the correlations between norm perceptions and punishment do not substantially change after including the own transfer.

Table C.3: Tobit regression punishment on deviations of norm perceptions from transfers

	Dependent Variable:			
	Punishment			
	(1)	(2)	(3)	(4)
Dev Personal Norms	0.26***			0.23**
	(0.04)			(0.07)
Dev Normative Expect.		0.25***		−0.27*
		(0.04)		(0.11)
Dev Empirical Expect.			0.31***	0.34**
			(0.04)	(0.11)
Constant	−0.46	−0.12	0.91	0.97
	(2.81)	(2.87)	(2.76)	(2.70)
Neg. Emotions	✓	✓	✓	✓
Δ Neg. Emotions	✓	✓	✓	✓
Observations	888	888	888	888
# Clusters	296	296	296	296
Log Likelihood	−2,434.61	−2,451.65	−2,433.95	−2,420.52
Wald Test	94.68***	62.68***	97.52***	122.10***
	(df = 3)	(df = 3)	(df = 3)	(df = 5)

Note: SE clustered at individual level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

C.1.4 Interaction Empirical Expectations and Strategy Method

In this section, we check whether subjects decrease their punishment for a transfer in the strategy method, which they believe is more likely to occur. We declare the dummy variables Neighborhood5, Neighborhood15, and Neighborhood25, which take the value 1 if the to-be-punished transfer in the strategy method is within the distance of 5 CZK, 15 CZK, or 25 CZK from a subject's empirical expectations, respectively. Table C.5 shows Tobit regressions, which include norm perceptions and neighborhood dummies. Models (4) and (5) include an interaction of the neighborhood variable and the Transfer dummies. We allow for such interaction because the effect of the neighborhood variable likely differs depending on the to-be-punished transfer. Specifically, the decrease in punishment may be less pronounced for a to-be-punished transfer of 40 because the initial punishment level is lower already compared to 0 or 10.

All models show a negative association between neighborhood variables and punishment decisions, but only at a significant level in model (2). This indicates that subjects seem to slightly decrease their punishment in the proximity of their empirical expectations. However, the neighborhood variable may simply capture higher punishment of a transfer of 0, 10,

Table C.4: Tobit regression punishment on own transfer in Section B and norm perceptions

	Dependent Variable:	
	Punishment	
	(1)	(2)
Transfer Section B	−0.04	−0.04
	(0.05)	(0.05)
Personal Norm		0.23**
		(0.07)
Normative Expect.		−0.24*
		(0.12)
Empirical Expect.		0.36**
		(0.11)
Constant	7.74**	0.32
	(2.94)	(3.57)
Transfer	✓	✓
Neg. Emotions	✓	✓
Δ Neg. Emotions	✓	✓
Observations	888	888
# Clusters	296	296
Log Likelihood	−2,456.35	−2,418.92
Wald Test	53.52***	125.08***
	(df = 5)	(df = 8)

Note: SE clustered at individual level $+$ $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

especially when not including the interaction with the Transfer dummies.

Most importantly, the positive significant association between personal norms and empirical expectations and punishment and the negative association with normative expectations (the results from the main text) remain robust and significant in all models. Hence, even though subjects might punish less severely in close proximity to their empirical expectations, empirical expectations still significantly explain punishment behavior.

Table C.5: Tobit regression punishment on norm perceptions and neighborhood dummy

	Dependent Variable:				
	Punishment				
	(1)	(2)	(3)	(4)	(5)
Neighborhood5	−1.92			−2.33	
	(1.40)			(3.78)	
Neighborhood15		-2.25^{*}			−0.42
		(1.05)			(3.04)
Neighborhood25			−1.48		
			(1.02)		
Personal Norms	0.23^{**}	0.23^{**}	0.24^{**}	0.23^{**}	0.23^{**}
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Normative Expect.	-0.25^{*}	-0.25^{*}	-0.25^{*}	-0.23^{+}	-0.23^{+}
	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
Empirical Expect.	0.33^{**}	0.31^{**}	0.32^{**}	0.37^{**}	0.43^{**}
	(0.11)	(0.11)	(0.11)	(0.12)	(0.14)
Constant	1.29	2.25	1.96	0.24	−1.01
	(2.39)	(2.39)	(2.39)	(2.61)	(3.59)
Transfer	✓	✓	✓	✓	✓
Transfer:Neighborhood5	×	×	×	✓	×
Transfer:Neighborhood15	×	×	×	×	✓
Observations	888	888	888	888	888
# Clusters	296	296	296	296	296
Log Likelihood	−2,423.00	−2,422.38	−2,423.07	−2,420.82	−2,420.64
Wald Test	118.11^{***}	119.65^{***}	118.20^{***}	122.85^{***}	123.05^{***}
	(df = 6)	(df = 6)	(df = 6)	(df = 8)	(df = 8)

Note: SE clustered at individual level. $+\ p < 0.1$; $^{*}\ p < 0.05$; $^{**}\ p < 0.01$; $^{***}\ p < 0.001$

C.2 Normative Expectations Gaps and Punishment

In this section, we run regressions, which estimate the correlations between third-party punishment and the differences between all pair-wise norm perceptions. Table C.6 shows the results of these regressions. In model (1), we find a significant positive association between punishment and the difference between personal norms and normative expectations. In model (2), we find a negative association between punishment and the difference between normative and empirical expectations. In model (3), we do not find any significant association between the difference between personal norms and empirical expectations with punishment. Finally, in model (4), we include both gaps PN-NE and NE-EE to analyze which relationship is more prevalent and stable. We find that the relationship of the gap NE-EE with punishment diminishes, while the significant positive association between punishment and the difference between personal norms and normative expectations prevails.

Table C.6: Tobit regression punishment on normative expectation gaps

	Dependent Variable:			
	Punishment			
	(1)	(2)	(3)	(4)
Gap PN-NE	0.26***			0.24**
	(0.07)			(0.07)
Gap NE-EE		−0.24*		−0.19+
		(0.11)		(0.11)
Gap PN-EE			0.10	
			(0.07)	
Constant	7.41*	8.86**	7.02*	8.42**
	(2.90)	(2.91)	(2.96)	(2.85)
Transfer	✓	✓	✓	✓
Neg. Emotions	✓	✓	✓	✓
Δ Neg. Emotions	✓	✓	✓	✓
Observations	888	888	888	888
# Clusters	296	296	296	296
Log Likelihood	−2,441.13	−2,449.52	−2,454.08	−2,436.64
Wald Test	82.48***	66.68***	57.87***	91.11***

Note: SE clustered at individual level + $p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

C.3 Causality

In this section, we provide the details of our analyses concerning the causal effect of social norm perceptions on third-party punishment. We first provide details on our analyses for reverse causality, then describe the IV analysis, and lastly provide details on our analysis for omitted-variable bias.

C.3.1 Reverse Causality

Table C.7 shows the differences between the first and second elicitation of norm perceptions in the Baseline. It reveals statistically significant higher personal norms and normative expectations in the second elicitation than in the first. Similarly, empirical expectations are higher in the second elicitation however not at a statistically significant level.

Table C.7: Comparisons of first and second norm perceptions elicitations in the Baseline treatment

Means (SE)	1st Elicitation	2nd Elicitation	p-value
Personal Norms	31.50 (2.36)	34.83 (2.33)	0.034
Normative Expectations	25.99 (1.59)	28.03 (1.66)	0.034
Empirical Expectations	22.22 (1.60)	23.25 (1.66)	0.915

Note: Non-parametric wilcoxon signed-rank test; N=60

Second, we compare the coefficients of the correlations between first and second norm perceptions elicitation. Table C.2 (Section C.1.3) replicates the findings of the main text. We find the same and significant relationships between all norm perceptions with punishment when looking at the first elicitation, i.e., the elicitations before the punishment (and experience) phase with a slightly larger negative coefficient for normative expectations and a lower positive coefficient for personal norms, but similar coefficients for empirical expectations. In addition, the coefficients (and significance levels) in the estimation of whether the punisher decided to punish (*punishment dummy*) are very similar between the first and second elicitation.

Third, we compare the first elicitation of norm perceptions between punishers and punishees. As Table C.8 depicts, there are no significant differences between the norm perceptions of punishers and punishees.

Table C.8: Comparisons of norm perceptions (first elicitation) between punishers and punishees

Means (SE)	Punishers *(N=296)*	Punishees *(N=119)*	p-value
Personal Norms	27.74 (1.15)	28.74 (1.85)	0.585
Normative Expectations	23.97 (0.70)	24.34 (1.19)	0.724
Empirical Expectations	19.53 (0.70)	20.66 (1.15)	0.492

Note: Non-parametric wilcoxon rank sum test

C.3.2 Instrumental Variable Approach

We employ a Tobit IV regression to estimate the causal influence of norm perceptions on punishment decisions. We use the Receiver and Observer treatments conditional on the transfers (Receive 0, ..., Receive 50, Observe 0, ..., Observe 50), as well as the assignment to the role of dictator (Dictator) as instruments for the second elicitation of each norm perception separately as well as for negative emotions. Specifically, we use the following model specifications:

First stage:

$$y_i = treatments_i \Pi_1 + transfer_i \Pi_2 + \nu_i \tag{C.1}$$

Second stage:

$$punishment_i^* = y_i \beta + transfer_i \delta + \epsilon_i \tag{C.2}$$

As punishment is restricted in between 0 and 50, we do not observe *punishment*, but only:

$$punishment_i^* = \begin{cases} 0 & \text{if } punishment_i < 0 \\ punishment_i^* & \text{if } 0 \leq punishment_i \leq 50 \\ 50 & \text{if } punishment_i > 50 \end{cases} \tag{C.3}$$

Note that $punishment_i^*$ denotes one single punishment decision for a transfer of either 0, 10, or 40 of one subject. We cluster standard errors on the individual level to account for within-individual dependencies. $transfer_i$ is a vector of dummies for the (exogenous) transfers of 10, and 40. $treatments_i$ are the instruments (Receive 0, ..., Receive 50, Observe 0, ..., Observe 50, Dictator). y_i is a vector of the endogenous variables, i.e., personal norms, empirical expectations, normative expectations, negative emotions, and the difference in negative emotions between the first and the second elicitation. Further note that (ν_i, ϵ_i) are assumed to be distributed multivariate normal. Therefore, the first and second stages are estimated together by Maximum Likelihood.[1]

The treatments and transfers are exogenous to the punisher by design and thus fulfill the requirement of instrument exogeneity. To evaluate instrument relevance, we analyze how the instruments shift subjects' norm perceptions. For this, we compare the second elicitation of norm perceptions in the treatments to the Baseline. We show this change by regressing the norm perceptions on the instruments (see the linear regression in Table C.9). Note that the first stage in the Tobit IV is estimated simultaneously with the second stage via Maximum Likelihood. We show this linear regression for illustrative purposes only.

In line with the literature on the erosion of social norms (Bicchieri et al., 2022; Keizer et al., 2008; Gino et al., 2009), we find that a norm violation, i.e., a transfer of 0 and 10, has a stronger effect on shifting social norm perceptions compared to high transfers: A low transfer (0, 10) decreases all three types of norm perceptions compared to the Baseline. The effect of a high (40, 50) transfer on norm perceptions is ambiguous and depends on receiving

[1] See `https://www.stata.com/manuals/rivtobit.pdf` for more information on the estimation procedure.

Table C.9: Linear regression norm perceptions and emotions on instruments

	Dependent Variable:				
	Pers. Norm	*Norm. Exp.*	*Emp. Exp.*	*Neg. Em.*	*Δ neg. Em.*
	(1)	(2)	(3)	(4)	(5)
Dictator	-9.46^{**}	-6.47^{**}	-4.19^{+}	-0.11	-0.22^{*}
	(3.22)	(2.17)	(2.28)	(0.18)	(0.11)
Receive 0	-2.61	-5.85^{+}	-14.40^{***}	1.05^{***}	0.83^{***}
	(4.61)	(3.21)	(2.28)	(0.29)	(0.20)
Receive 10	-6.20	-6.37^{*}	-6.84^{**}	0.89^{**}	0.21
	(5.08)	(3.03)	(2.55)	(0.28)	(0.17)
Receive 40	3.99	3.58	7.18^{**}	0.82^{**}	0.06
	(4.03)	(2.83)	(2.73)	(0.30)	(0.15)
Receive 50	1.53	4.61	-3.12	-0.08	0.07
	(5.92)	(4.19)	(3.23)	(0.29)	(0.16)
Observe 0	-8.54^{+}	-7.51^{*}	-6.94^{*}	0.27	0.01
	(4.51)	(2.99)	(2.78)	(0.26)	(0.12)
Observe 10	-9.42^{*}	-5.12^{+}	-3.27	-0.01	0.19
	(4.50)	(3.01)	(2.92)	(0.25)	(0.16)
Observe 40	-2.48	-0.52	2.00	-0.49^{+}	-0.10
	(4.95)	(3.57)	(3.53)	(0.27)	(0.12)
Observe 50	-10.29	-2.09	-6.57	0.12	0.02
	(6.62)	(3.92)	(4.16)	(0.39)	(0.13)
Constant	34.83^{***}	28.03^{***}	23.25^{***}	2.47^{***}	0.02
	(2.36)	(1.67)	(1.68)	(0.14)	(0.07)
Observations	296	296	296	296	296
R^2	0.06	0.08	0.15	0.14	0.16
Adjusted R^2	0.03	0.05	0.12	0.12	0.13
Residual Std. Error	19.13	12.54	11.97	1.08	0.67
F Statistic (df = 9; 286)	1.93^{*}	2.84^{**}	5.50^{***}	5.37^{***}	5.85^{***}
Note:	$+p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$				

or observing a transfer. Even though the Observer and Receiver treatments give a signal of what is typically done, the instruments do not only significantly shift empirical expectations but also normative expectations - however, to a smaller extent. Personal norms also get shifted, yet mostly not significantly. Being assigned to the role of the dictator significantly shifts personal norms, normative expectations, and empirical expectations. We conclude that the instruments significantly shift all three norm perceptions and hence are relevant.

Finally, a valid instrument has to fulfill the exclusion restriction property. The instruments should exclusively influence punishment through the instrumented variables. To ensure this, we include the channel of negative emotions, as they are known to correlate with punishment (Jordan et al., 2016a; Carpenter and Matthews, 2012; Nelissen and Zeelenberg, 2009). We confirm this by finding a (marginally) significant correlation between the difference between negative emotions and punishment (see Table 3.1). Additionally, we observe that the instruments influence negative emotions and the difference between negative emotions in the case of receiving a transfer or being assigned to the role of a dictator. Hence, we include both negative emotions and the difference between negative emotions in the IV regression in the main text (Table 3.2).[2]

Table C.10 replicates the results reported in the main text without instrumenting for negative emotions and the change in negative emotions. We find a positive influence of each of the norm perceptions individually.

C.3.3 Omitted-Variable Bias

In this section, we run an additional robustness check to tackle a potential omitted-variable bias. For doing so, we include the following individual-level controls to the regression: age, gender, income, field of study,[3] degree of understanding, and degree of concentration. As an additional robustness check, we omit negative emotions. Table C.11 shows these regressions. Model (1) depicts the model without negative emotions, model (2) includes negative emotions, and finally, model (3) includes the individual-level controls. All models indicate the same correlations between social norm perceptions with punishment. The negative correlation of normative expectations with punishment, yet, becomes only marginally significant. Nonetheless, the robustness check replicates the relationships, giving further evidence for a causal influence of social norm perceptions on third-party punishment.

[2]The second elicitations incorporate the total change of negative emotions compared to the Baseline, however only on an aggregate level. The differences between the first and second elicitations incorporate individual changes. This is particularly important for the self-report of negative emotions because subjects may interpret the 7-Likert scale differently from each other. By focusing on the change of negative emotions, those individual differences in the absolute interpretation of the scale get less pronounced.

[3]The base field of study is 'Others'.

Table C.10: Tobit IV regression punishment on norm perceptions, second stage

	Dependent Variable:			
	Punishment			
	(1)	(2)	(3)	(4)
Personal Norms	0.58^{+} (0.34)			0.15 (1.00)
Normative Exp.		0.76^{*} (0.36)		0.35 (1.45)
Empirical Exp.			0.50^{*} (0.25)	0.26 (0.45)
Transfer10	-4.24^{***} (0.84)	-4.20^{***} (0.83)	-4.23^{***} (0.83)	-4.27^{***} (0.84)
Transfer40	-11.50^{***} (1.30)	-11.45^{***} (1.29)	-11.45^{***} (1.28)	-11.55^{***} (1.29)
Constant	-8.78 (10.18)	-10.12 (8.83)	-1.41 (4.85)	-9.67 (8.35)
Observations	888	888	888	888
Log Likelihood	-6299.94	-5939.23	-5885.93	-12771.59
Wald χ^2 (df = 3)	84.89^{***}	87.38^{***}	86.58^{***}	89.30^{***}

Note: SE clustered at individual level. $+\ p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Treatments conditional on transfer as instruments for each social norm perception separately.

Table C.11: Tobit regression punishment on social norm perceptions and controls

	Dependent Variable:		
	Punishment		
	(1)	(2)	(3)
Personal Norm	0.23^{**}	0.23^{**}	0.22^{**}
	(0.07)	(0.07)	(0.07)
Normative Expect.	-0.25^{*}	-0.25^{*}	-0.20^{+}
	(0.12)	(0.12)	(0.12)
Empirical Expect.	0.34^{**}	0.36^{**}	0.35^{**}
	(0.11)	(0.11)	(0.11)
Neg. Emotions		0.05	0.04
		(0.94)	(0.90)
Δ Neg. Emotions		2.45^{+}	2.16^{+}
		(1.26)	(1.20)
Age			0.57
			(0.56)
Female			−1.55
			(2.15)
Income			0.0000
			(0.0001)
Highest Degree			−0.93
			(2.04)
Economics/ Business			2.19
			(4.07)
Engineering/ IT			7.02
			(5.92)
Humanities/ Medicine/ Education			5.34
			(4.35)
Natural Sciences			8.17
			(5.31)
Social Sciences			7.55
			(6.20)
Degree of Understanding			-2.55^{*}
			(1.13)
Degree of Concentration			0.37
			(1.15)
Constant	0.82	0.29	−5.71
	(2.41)	(3.59)	(12.00)
Transfer	✓	✓	✓
Observations	888	888	888
# Clusters	296	296	296
Log Likelihood	−2,423.60	−2,419.70	−2,405.91
Wald Test	116.63^{***} (df = 5)	123.80^{***} (df = 7)	150.21^{***} (df = 18)

Note: SE clustered at individual level $+p < 0.1; *p < 0.05; **p < 0.01; ***p < 0.001$

C.4 Heterogeneity in Norm-driven Punishment

In this section, we study how punishment and the relative importance of the three norm perceptions may differ between subjects. For this, we first explore the impact of the exogenous assignment to the different roles (treatments) in the Experience Phase. Additionally, to illustrate subject-specific heterogeneities, we study how norm-driven punishment decisions change with gender.

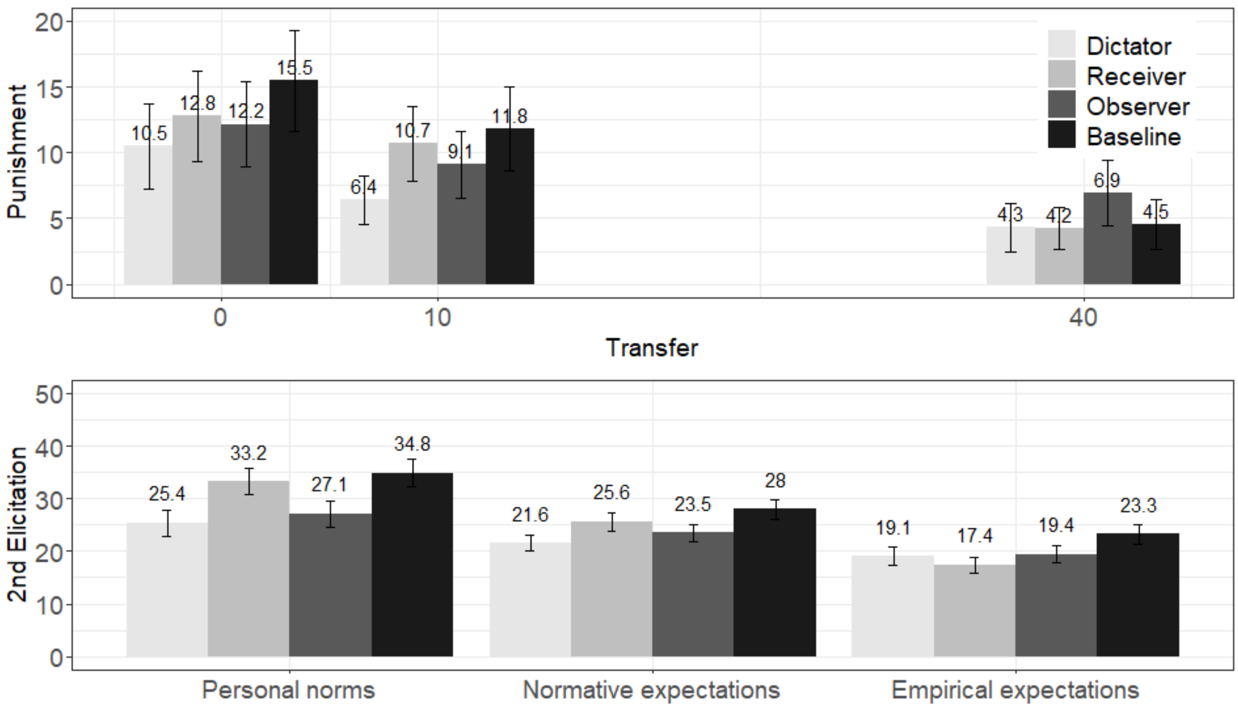
C.4.1 Roles and Norm-driven Punishment

Figure C.4 shows punishment decisions conditional on the role in the Experience Phase and their respective norm perception averages. Non-parametric Wilcoxon rank sum tests reveal that the mere assignment to the role of dictators in the Experience Phase significantly (at least $p < 0.05$) reduces punishment for a transfer of 0 and 10 compared to the Baseline.

Importantly, the figure indicates that the changes in punishment decisions (for a transfer of 0 and 10) closely follow the changes in the norm perceptions induced by the treatments. For instance, in the Baseline, both punishment decisions and norm perceptions are the highest of all treatments. Similarly, in the Dictator treatment, both punishment decisions and personal norms and normative expectations are the lowest. This illustrates the importance of norm perceptions for punishment decisions, also if the differences are induced by the exogenous assignment to a specific treatment.

Furthermore, the lower panel of Figure C.4 reveals that the between-treatment differences in empirical expectations do not follow the same pattern as the differences in personal norms and more importantly, punishment decisions. This indicates that the importance of the three norm perceptions may differ depending on the role. To study whether this is the case, we regress punishment on the norm perceptions and interact them with the treatments (see Table C.12). We find that being in the Receiver treatment (marginally) significantly ($p < 0.1$) increases the importance of empirical expectations for punishment compared to all other treatments. Apart from this, there is no other significant interaction between the treatment assignment and one of the norm perceptions. We conclude that the exogenous assignment into a specific role may change the importance of empirical expectations, however, the relative importance seems to be rather stable for the different roles.

Figure C.4: Treatments, social norm perceptions, and punishment



Note: The upper panel shows punishment decisions conditional on the transfer of the to-be-punished dictator (Transfer) and on the specific role in the Experience Phase. The lower panel shows norm perception averages conditional on the specific role. Error bars show 95% confidence intervals and are based on standard errors that are clustered on individual level.

Table C.12: Tobit regression punishment on the role and interaction with norm perceptions

	Dependent Variable:			
	Punishment			
	(1)	(2)	(3)	(4)
Personal Norm	0.20*	0.22*	0.24**	0.25**
	(0.09)	(0.09)	(0.08)	(0.08)
Normative Expect.	−0.19	−0.18	−0.28*	−0.31*
	(0.14)	(0.15)	(0.14)	(0.13)
Empirical Expect.	0.39**	0.22	0.38**	0.41***
	(0.13)	(0.14)	(0.14)	(0.12)
Dictator	3.06			
	(5.20)			
Receiver		−8.22		
		(5.50)		
Observer			0.95	
			(5.36)	
Baseline				5.62
				(7.25)
Personal Norm:Treatment	0.08	0.06	−0.01	−0.16
	(0.15)	(0.14)	(0.18)	(0.21)
Normative Expect.:Treatment	−0.37	−0.14	0.15	0.34
	(0.29)	(0.25)	(0.27)	(0.31)
Empirical Expect.:Treatment	−0.002	0.46+	−0.12	−0.30
	(0.25)	(0.25)	(0.25)	(0.32)
Constant	0.14	1.86	−0.23	−0.64
	(4.06)	(3.68)	(3.82)	(3.94)
Interaction:Treatment	Dictator	Receiver	Observer	Baseline
Transfer	✓	✓	✓	✓
Neg. Emotions	✓	✓	✓	✓
Δ Neg. Emotions	✓	✓	✓	✓
Observations	888	888	888	888
# Clusters	296	296	296	296
Log Likelihood	−2,414.39	−2,413.75	−2,418.54	−2,415.53
Wald Test (df = 11)	134.16***	135.12***	125.97***	131.48***

Note: SE clustered at individual level $+\ p < 0.1$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

C.4.2 Gender

In this section, we look closer at gender differences in the importance of norm perceptions for punishment decisions. In the existing literature, the results about the effect of gender on punishment remain mixed. Kromer and Bahçekapili (2010) find that males punish selfish behavior more often than females and McAuliffe et al. (2015) confirm this results among children. In contrast, Carpenter and Matthews (2012) find that females punish more than males and Leibbrandt and López-Pérez (2012) show that females engage in more antisocial punishment. Piardini et al. (2017) study different gender compositions of punisher and punishee and find that males punish females significantly more than females punish males and that same-sex groups do not differ in punishment. Moreover, there is only little evidence on how norm perceptions are related to economic behavior with respect to gender. Croson et al. (2010) find that males rely more on their empirical expectations when deciding about their own donations. Krysowski and Tremewan (2021) find that females find giving unfair amounts in a dictator game less acceptable compared to males when the dictator is unidentified. Fišar et al. (2016) study gender differences in bribing behavior and do not find a gender difference in the positive association between accepting bribes and beliefs about how often others accept bribes.

In our experiment, we do not find significant differences in third-party punishment decisions between males and females. However, we find differences in the relative importance of norm perceptions for punishment decisions across genders. For this, we first split the sample into males and females. Table C.13 shows the results of a Tobit model, where we regress punishment on norm perceptions. Model (1) includes only females, model (2) only males, and model (3) the full sample with a dummy indicating females and the interaction of females with all norm perceptions. In model (3), the interaction effect shows the importance of norm perceptions for punishment for females, whereas the baseline effect shows this relationship for males. In model (1), we find a positive relationship between empirical expectations and punishment for females and a negative relationship with normative expectations. Personal norms seem not to matter for females. In model (2), we find the opposite for males. We find a statistically significant positive association between personal norms and punishment, whereas empirical and normative expectations seem not to matter for their punishment decisions. Model (3) confirms this pattern, as there is a significant negative interaction effect between females and personal norms and a significant positive interaction effect between females and empirical expectations.

We find that this difference in the importance of the norms for punishment decisions is not driven by different initial (first elicitation) levels of norm perceptions between females

Table C.13: Tobit regression punishment on norm perceptions and gender

	Dependent Variable:		
	Punishment		
	(1)	(2)	(3)
Personal Norm	0.08	0.48^{***}	0.45^{***}
	(0.09)	(0.13)	(0.12)
Normative Expect.	-0.29^{*}	-0.22	-0.21
	(0.15)	(0.19)	(0.17)
Empirical Expect.	0.55^{***}	0.05	0.06
	(0.14)	(0.18)	(0.18)
Female			2.95
			(4.76)
Female:Personal Norm			-0.36^{*}
			(0.15)
Female:Normative Expect.			-0.09
			(0.23)
Female:Empirical Expect.			0.49^{*}
			(0.22)
Neg. Emotions	-0.32	0.24	0.04
	(1.14)	(1.70)	(0.94)
Δ Neg. Emotions	4.39^{*}	0.73	2.33^{+}
	(2.00)	(1.44)	(1.20)
Constant	1.67	-0.75	-1.03
	(4.46)	(5.73)	(4.32)
Gender	Female	Male	Both
Transfer	✓	✓	✓
Observations	489	399	888
# Clusters	163	133	296
Log Likelihood	$-1{,}337.91$	$-1{,}063.89$	$-2{,}406.57$
Wald Test	72.39^{***} (df = 7)	77.82^{***} (df = 7)	147.31^{***} (df = 11)

Note: SE clustered at individual level. $+\ p < 0.1$; $*\ p < 0.05$; $**\ p < 0.01$; $***\ p < 0.001$

and males. Table C.14 shows average norm perceptions for females and males. The only substantive difference in norm perceptions is between empirical expectations (20.79 for males vs. 18.57 for females), yet the difference is not statistically significant.

Table C.14: Differences first elicitation norm perceptions between gender

Means (SE)	Males *(N=163)*	Females *(N=133)*	p-value
Personal Norms	30.00 (1.70)	29.63 (1.51)	0.922
Normative Expectations	24.85 (1.17)	24.11 (0.97)	0.668
Empirical Expectations	20.79 (1.14)	18.57 (0.97)	0.145

Note: Non-parametric Wilcoxon rank sum test

We conclude that females want to enforce typical behavior, whereas males punish according to what they personally believe is appropriate. This result illustrates that there exist heterogeneities in the importance of each of the norm perceptions for third-party punishment.

C.5 Experimental Instructions

The experiment took place online with the subject pool of the Masaryk University Experimental Economics Laboratory (MUEEL). We enclose the experimental protocol and instructions with experimental screens for punishers and punishees. The subjects went through pages independently, and the experimenter was present at the Zoom meeting, communicating with participants through the Zoom chat. If some participants did not show any activity for more than 2 minutes (apart from the planned waiting time within instructions), the experimenter contacted them through the chat or called them on the phone in case they did not respond.

Protocol and Verbal Instructions for Online Sessions for EXPERIMENTER

Main Experimenter - checks in participants, gives all verbal instructions + number, runs ZTU from VM, sends them the individual links to participants, solves ZTU issues if they come up, handles private chat if necessary.

20 - 30 MINUTES BEFORE SESSION BEGINS:

<Host/ Experimenter 1 checks in participants one by one, once their audio connects, says>

EXPERIMENTER 1: "Can you hear me? If you can, please unmute yourself and let me know. < if full name is not clear from zoom name, ask for it – Can you tell me your full name? – thank you>. ***Now I will assign you the number, which we will use at the beginning of the experiment. Your number is X.*** *I will now direct you to a breakout room where you should wait for the experiment to begin. You don't have to be at computer now, just be fully prepared at least 5 minutes before the beginning of the experiment. Once you see join, please click on it, and you can mute yourself and turn the camera off now.*

<Direct participant to breakroom where on second computer shares screen (see Experiment_Lobby_Screen), >

Welcome to the experiment lobby! The experiment will begin shortly…

- The random number you got assigned is your participant label. We will use it in case we have more participants than we need in this experiment. If that is the case, we will randomly draw a number – if your number will be drawn, you won't participate in today's experiment, but we will pay you 50 CZK for your show up.
- Please turn off your video and put yourself on mute. It is not allowed to communicate with other participants for the whole duration of the experiment.
- Please keep **your unique ID (or code)** ready as we will ask you to enter it during this session. (You can find this in the email you received with the zoom link for this session).
- Make sure you are in a calm, quiet area without any distractions or people around. While the experiment is running, please stay focused. Please make sure to close all applications on your computer and any websites that are open.
- If you have any questions at any point, please type them into the private chat feature on Zoom and one of the experimenters will respond.
- You have to be fully prepared at least 5 minutes before the beginning of the experiment.

Thank you for your cooperation.

At the beginning - ASKING RESERVES TO LEAVE:

<Experimenter enters breakout room and inform participants, that it will be closed and they will get instructions in main meeting >

MAIN EXPERIMENTER: "Hello again and thanks for coming. For this experiment, we require 14 people. All of you have been assigned a number between 1 and X. I will now share my screen with random number generator. If your number is randomly generated, I will ask you to leave this zoom meeting. If you have entered your bank details in hroot then you will receive CZK 50 for this experiment, and you can register for another session of this experiment. Thanks for coming"

<repeat for all numbers, afterward makes sure that the reserves have left the zoom meeting room->

STARTING THE SESSION:

MAIN EXPERIMENTER: "I will now be sending you your individual links to the experiment. I need to do this one by one so please keep an eye on the chat and you should receive your link in a few minutes. After receiving the link, please paste it in your browser. The experiment will begin shortly after that. If you have any questions during the experiment, please use the chat feature on zoom to ask the question and we will respond to it there.>

<send individual links to all via private chat>

MAIN EXPERIMENTER: "You have now all received the link to the online experiment, so we can start soon. You should see grey screen with small green leaf on side, if you do not, please write to me through chat. If you have not done so already, please now minimize this zoom meeting (without closing it) and move it away from your screen. Since this is an interactive experiment, you might have to wait while other participants make decisions but it is important that you do not engage in any other activity during this time. Please do not open up any tabs on your browser. We will begin shortly."

<Experimenters go on mute, we make sure background is set to number of links that were sent out, press F5 in the VM and start >

ENDING THE SESSION

MAIN EXPERIMENTER: "Now you see payment screen. This is the last screen of this experiment. If you have any question or feedback, please, write to us. Thank you for participating in this experiment. Your full payment will be transferred to your accounts until the end of two working days. After you click on proceed, you can close the tab and leave the zoom meeting room. Thanks again and goodbye."

--DONE--

Instructions – punishers

[screen 1]

Experimental instructions

You are now taking part in an economic experiment. You can earn a considerable amount of money depending on the decisions that you and other participants make. Therefore, it is very important that you read the instructions carefully. It is important to us that you stay concentrated and in front of your computer. Communication with any of the participants is strictly forbidden and can lead to withholding of the payment.

This experiment consists of a part A and a part B. Either part A or part B is going to be paid out to you. Part A will be paid with a probability of 80% and part B with a probability of 20%. You will receive your payoff on your bank account within two working days from the end of the experiment.

If you have questions or technical problems, please write to us through the chat in zoom.

[screen 2 – dictators and receivers]

Experimental instructions - Part A - Stage 1

We will now explain part A. After reading the instructions for the entire part A, you will start to make decisions. Therefore, carefully read the instructions and if you have any questions, please write to us through the chat in zoom. Instructions for part B will be shown after you have finished part A. Part A consists of two stages.

In the first stage, you will be paired randomly and anonymously with another participant. One of you will be randomly assigned to be Player A and the other to be Player B.

Player A will receive 100 CZK and Player B will not receive anything. Player A can then decide to transfer either 0, 10, 40, or 50 CZK to Player B.

Your role will be Player A. [Your role will be Player B.]

[screen 3 – all treatments]

Experimental instructions – Part A [- Stage 2 – dictators and receivers]

[In stage 2, – dictators and receiver] There are players C and D, who are like you real human participants of this experimental session. You will make decisions that will affect the payoff of Player C, therefore your choices have real consequences for your own payoff and for the payoff of Player C.

Player C receives 100 CZK and Player D does not receive anything. Player C can then decide to transfer either 0, 10, 40, or 50 CZK to Player D.

You can assign deduction points to participant C. Each deduction point you transfer to participant C diminishes your income by 1 CZK and participant C's income by 3 CZK. You can assign a number of deduction points between 0 and 50. You will decide how many deduction points to assign to Player C for any possible choice of him/her. Specifically, you will decide how many deduction points to assign if Player C transfers either 0, 10, 40, or 50 CZK.

You won't know how Players C have decided until the end of the experiment. Your choice will be implemented and the number of deduction points you chose will be assigned to Player C and his income will be reduced accordingly, depending on Player C's chosen transfer. Therefore, all your choices potentially have a real impact on the payoff of Players C. Note that nobody has the opportunity to assign deduction points to you at any point in the experiment.

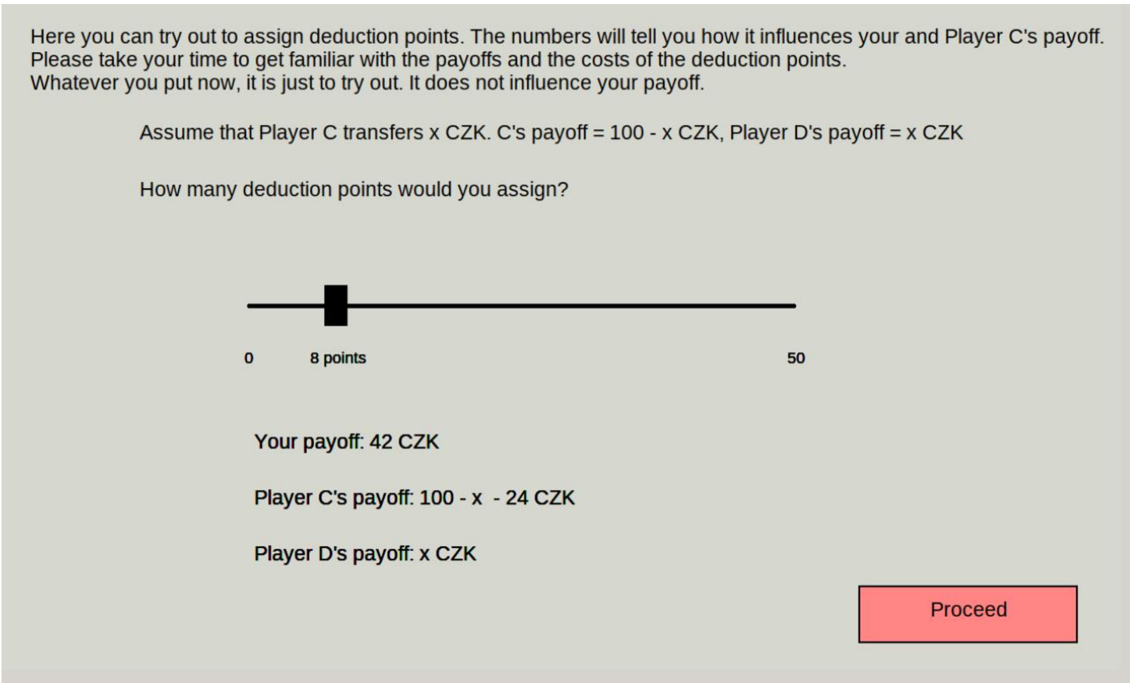
[screen 4 – tryout stage]

Here you can try out to assign deduction points. The numbers will tell you how it influences your and Player C's payoff. Please take your time to get familiar with the payoffs and the costs of the deduction points.

Whatever you put now, it is just to try out. It does not influence your payoff.

Assume that Player C transfers x CZK. C's payoff = 100 - x CZK, Player D's payoff = x CZK

How many deduction points would you assign?



[screen 5 – observers]

Before your decisions, we will give you an impression on how participants decided about the transfers. We will show you the decision and consequences of a randomly chosen participant, Player A, that participated in an earlier session of this experiment. That player is randomly chosen from all the players that participated in the earlier session and that were making a decision as Player A.

Player A received 100 CZK and Player B did not receive anything. Player A could then decide to transfer either 0, 10, 40, or 50 CZK to Player B. There was no opportunity to assign deduction points to Player A.

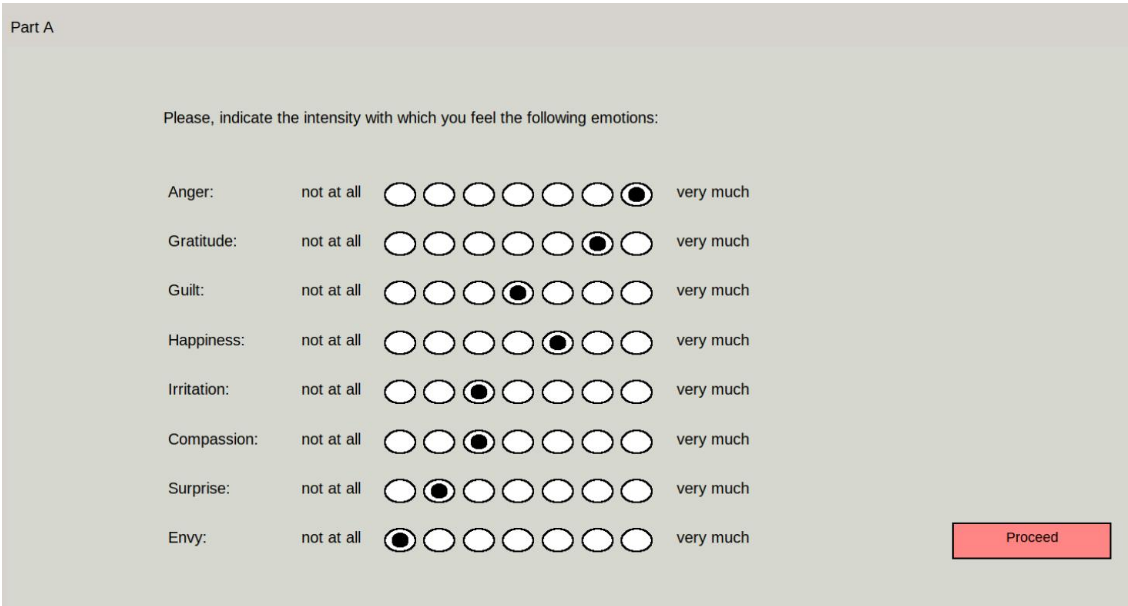
[screen 6 – all treatments]

Now we will start with part A. Remember that there is an 80% chance that this part is going to be payoff relevant for you. In this part A, you receive a base payment of 50 CZK [if observer or inactive 50 + X CZK; X randomly chosen payoff of subject from experience treatment in earlier session] that is independent of your future decisions and will be paid out for sure if part A is going to be picked for your payoff.

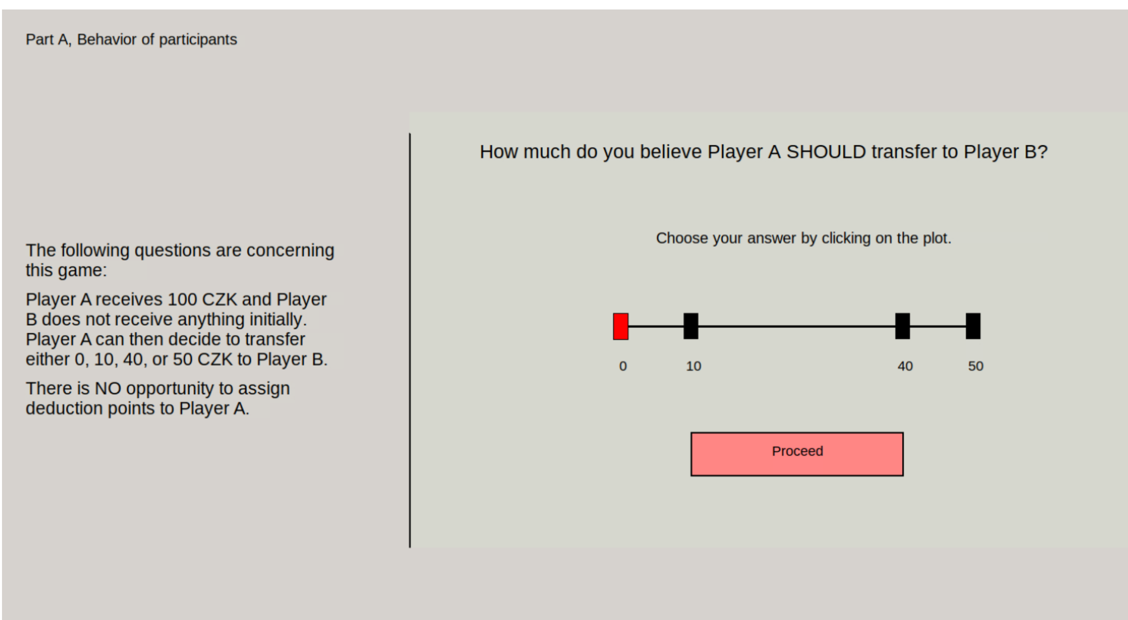
[Your role will be Player A [B] – experience]

Before we start with the first stage, we will ask you about your emotions and opinions on the behavior of participants in this [a previous – observe & inactive] experimental session.

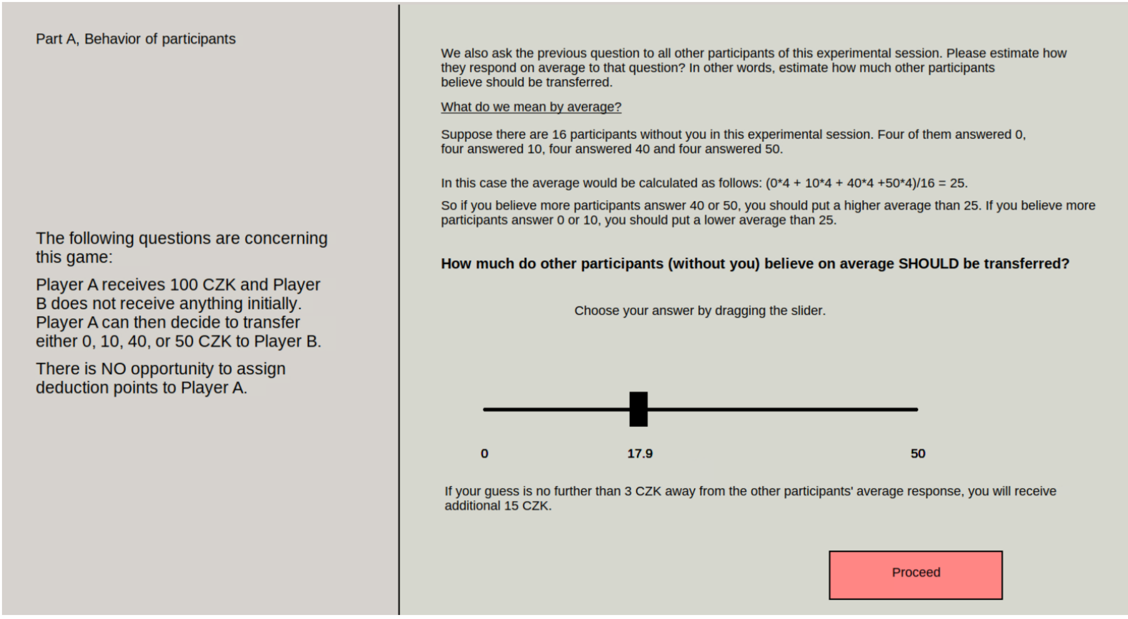
[screen 7 – emotions]



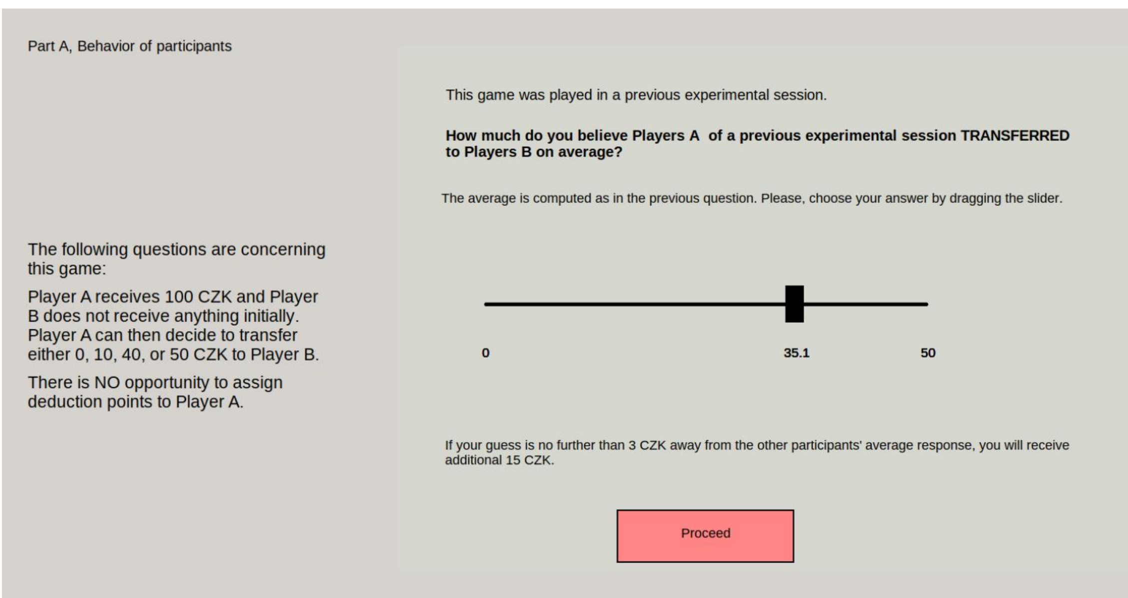
[screen 8 – norm elicitation – personal norm]



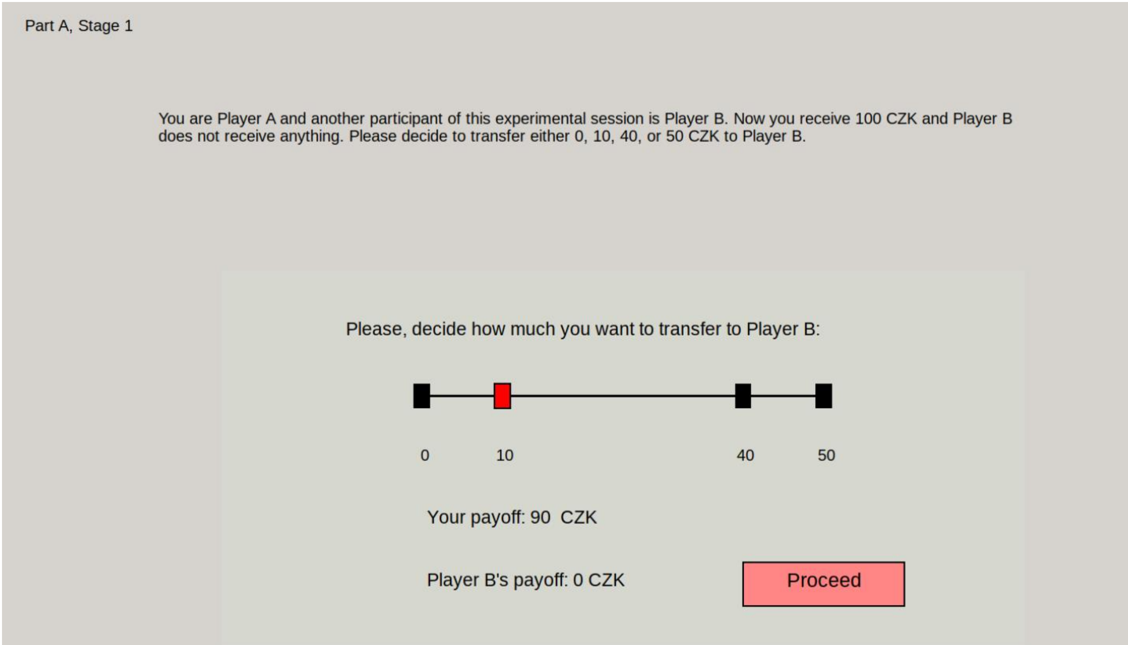
[screen 9 – norm elicitation – normative expectation]

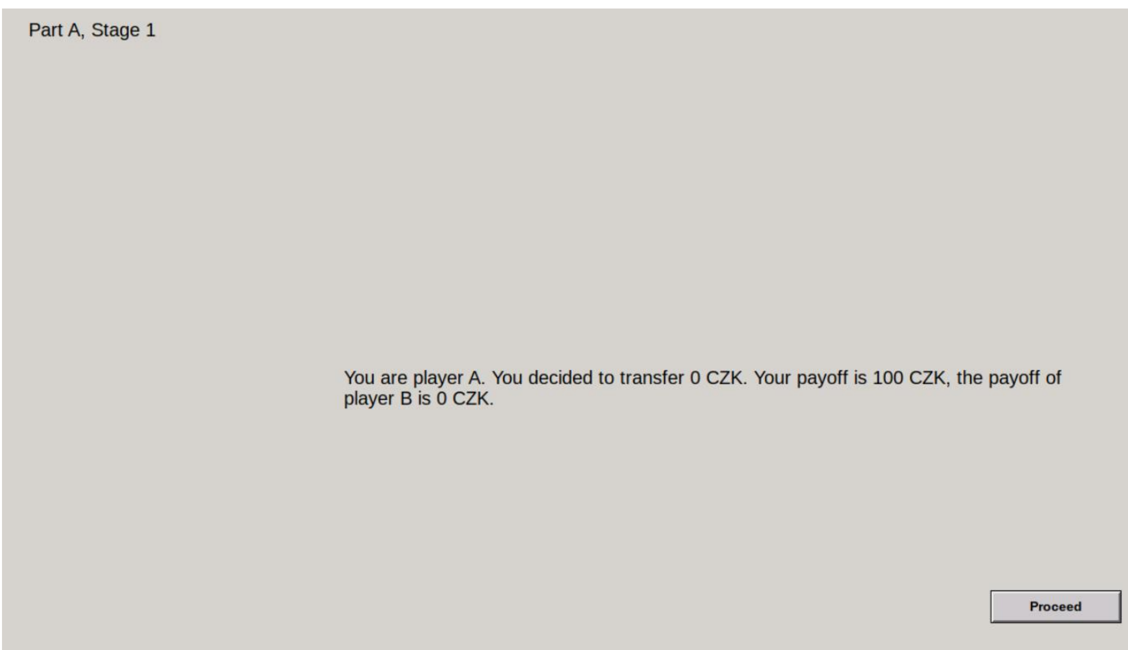


[screen 10 – norm elicitation – empirical expectation]

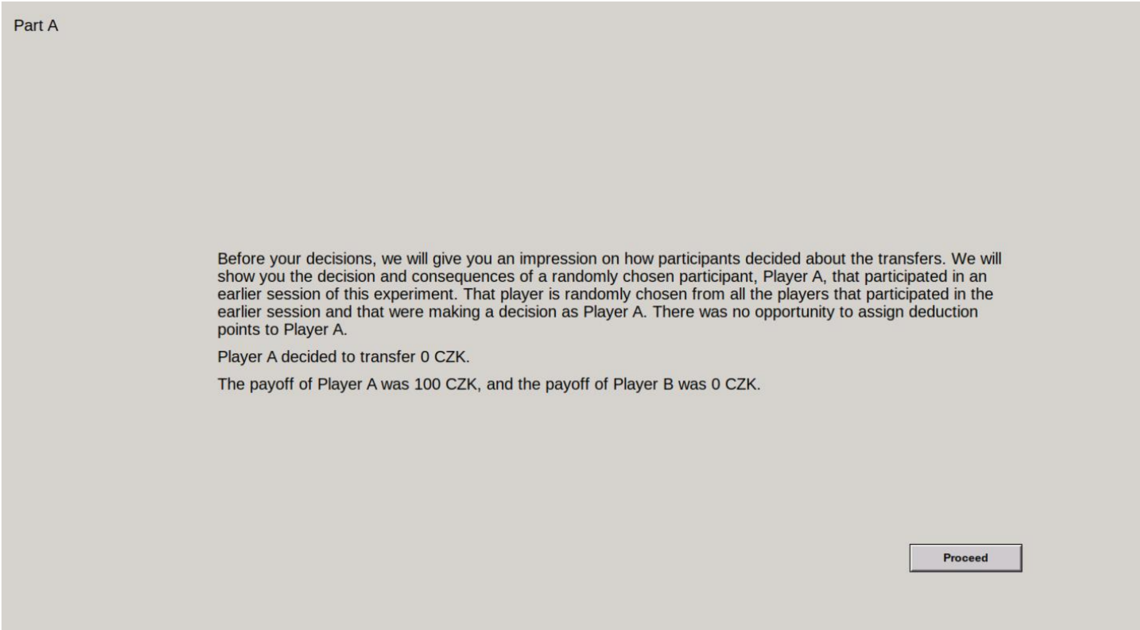


[screen 11 - Experience Phase – dictators]

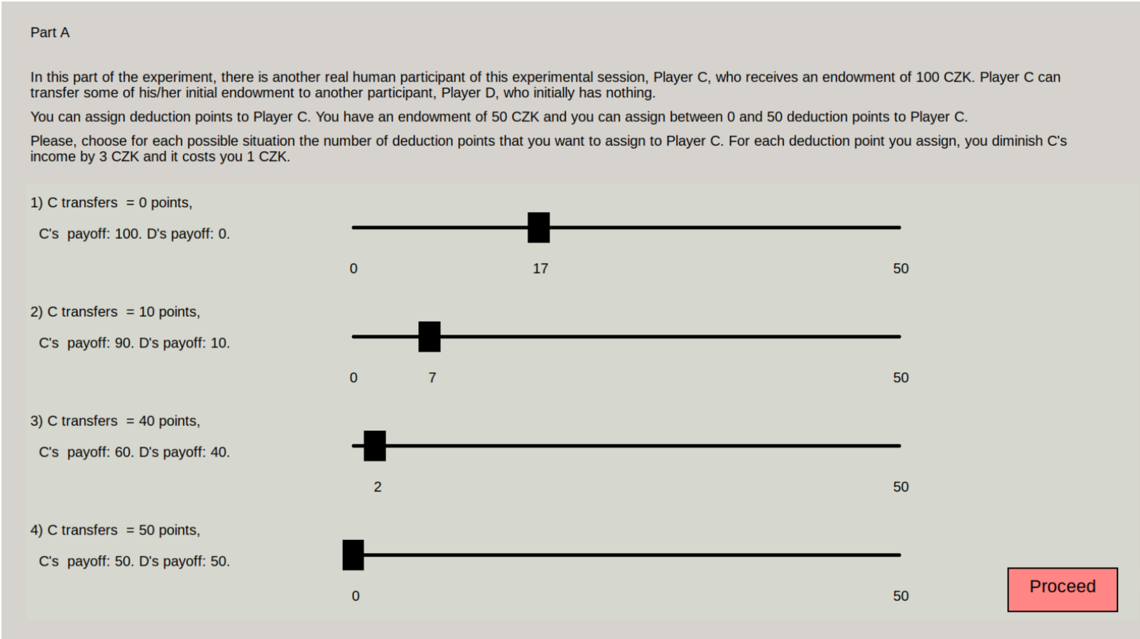




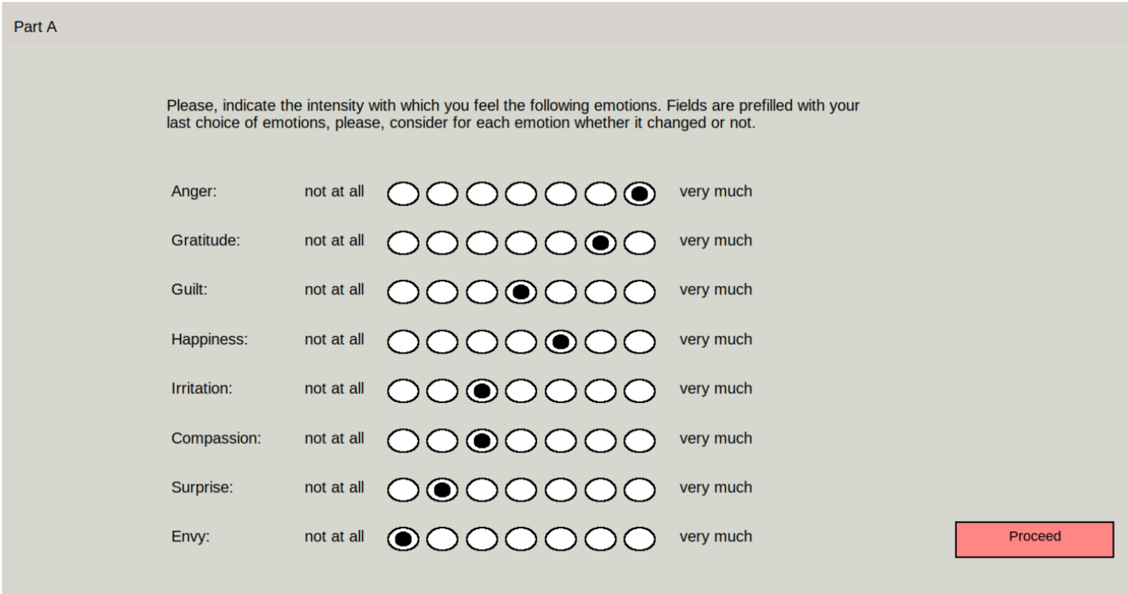
[screen 11 – Experience Phase - observers]



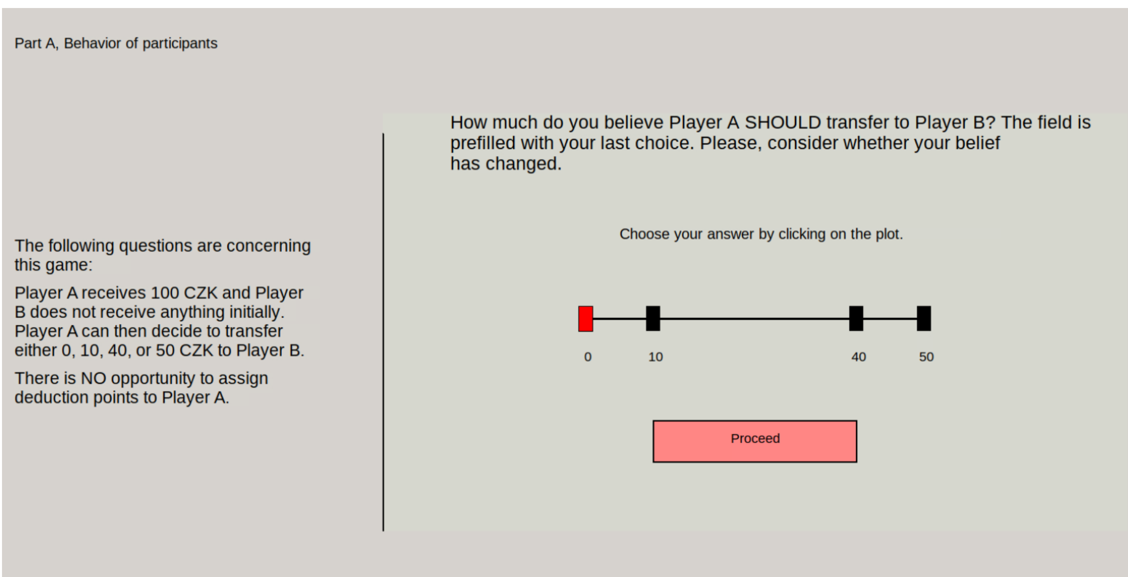
[screen 12 – Punishment Phase]



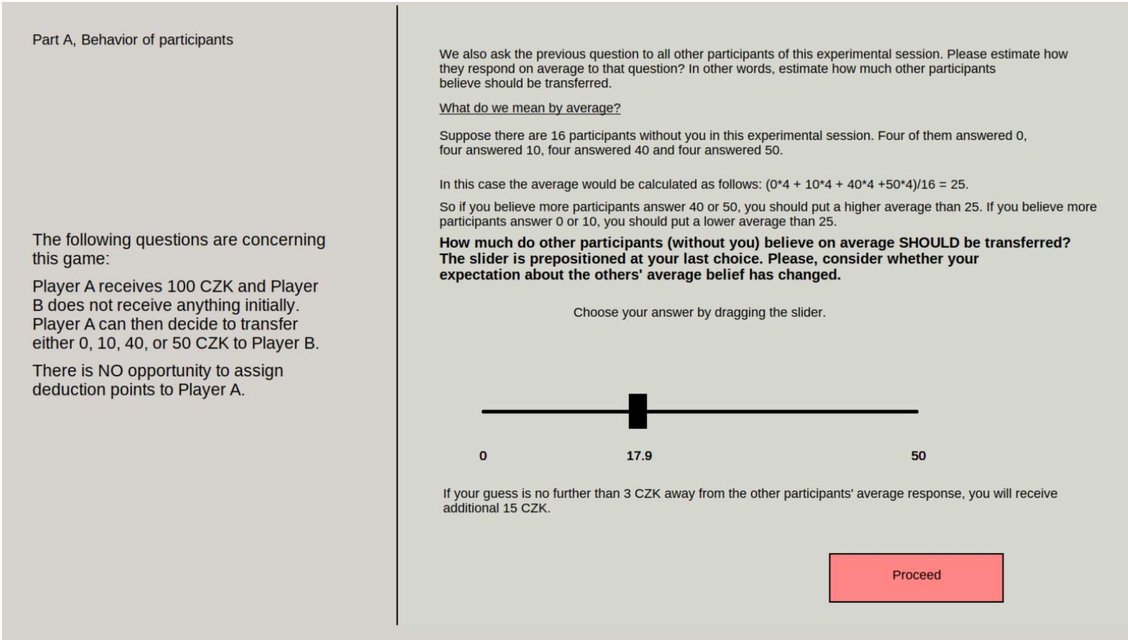
[screen 13 – emotions 2]



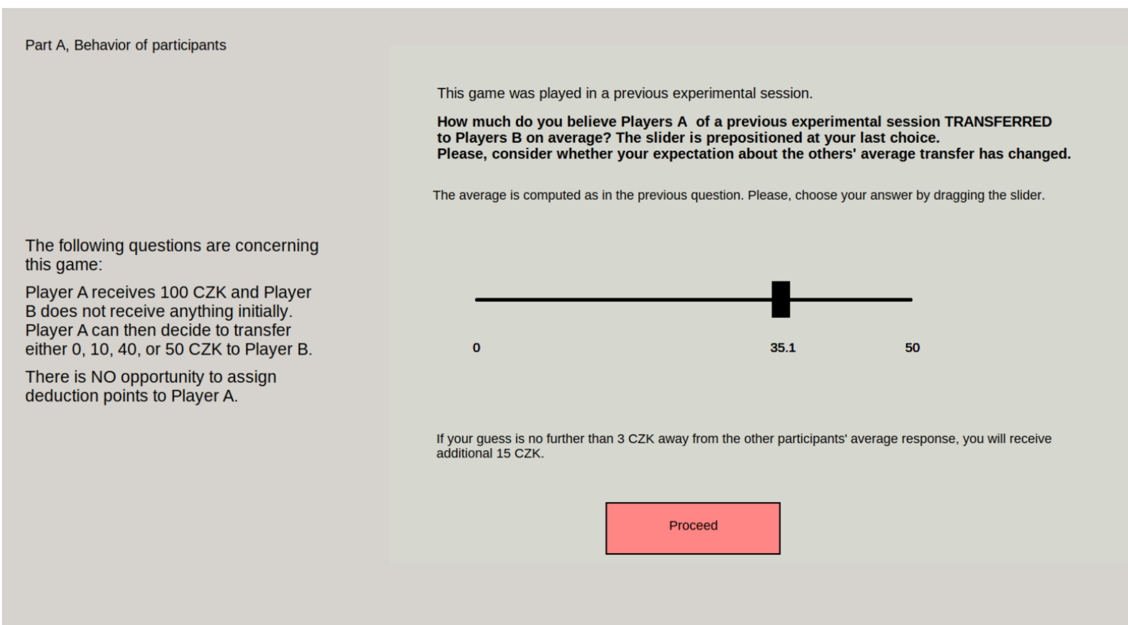
[screen 14 - norm elicitation 2 – personal norm]



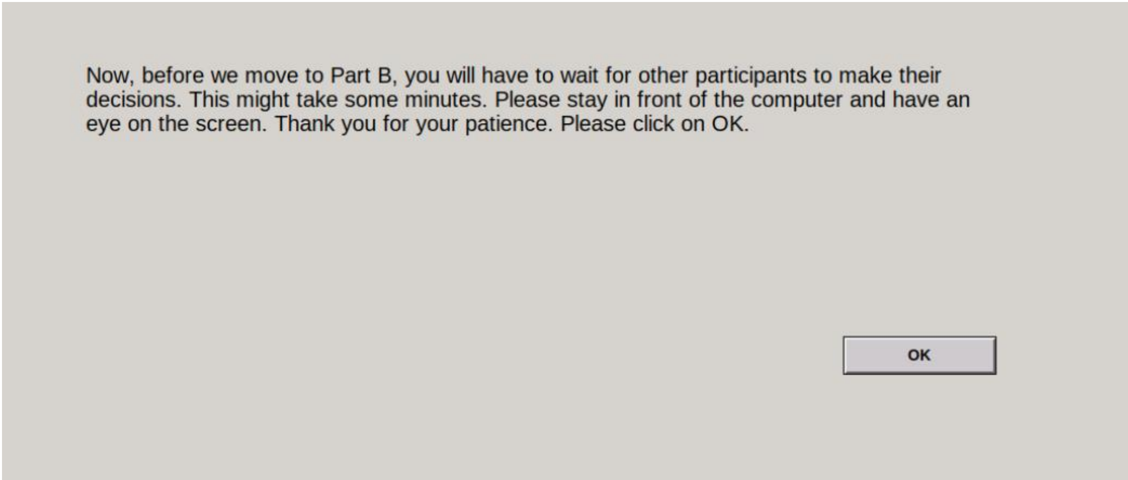
[screen 15 – norm elicitation 2 – normative expectation]



[screen 16 – norm elicitation 2 – empirical expectation]



[screen 17 - waiting stage]



[screen 18 – instructions part B]

Experimental instructions - Part B

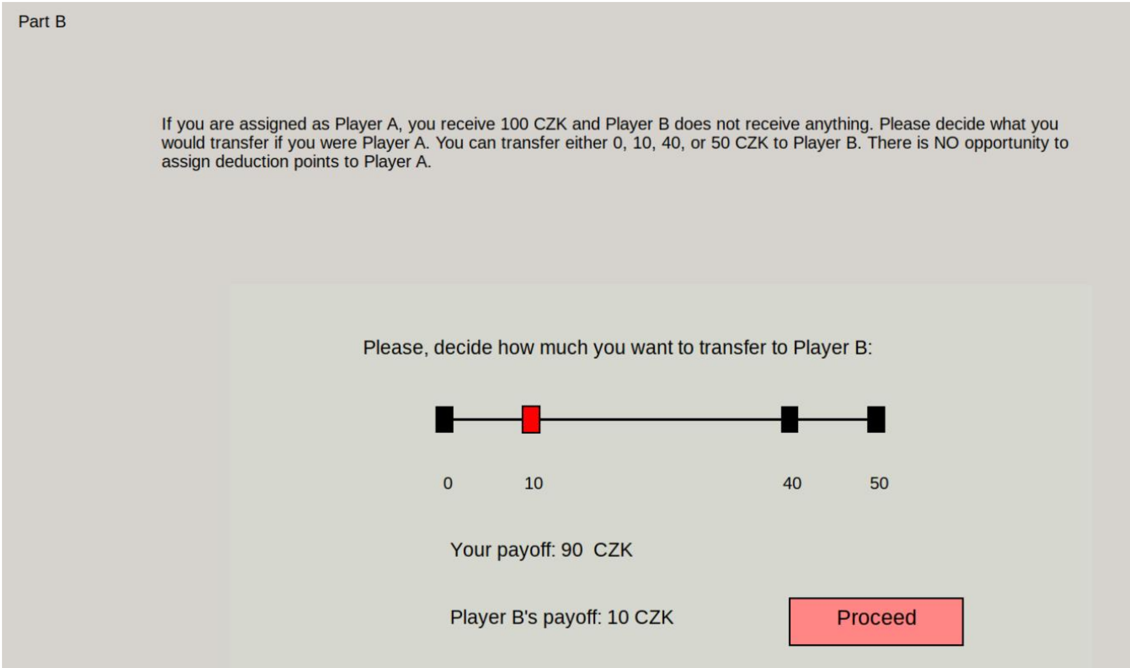
We have completed part A of this experiment. Now we will start with part B. In this part, you receive a base payment of 50 CZK that is independent of your future decisions and will be paid out for sure if part B is going to be picked for your payoff.

You will be paired with another participant of this experimental session. One of you will be randomly assigned to be Player A and the other to be Player B. Player A will receive 100 CZK and Player B will not receive anything. Player A can then decide to transfer either 0, 10, 40, or 50 CZK to Player B.

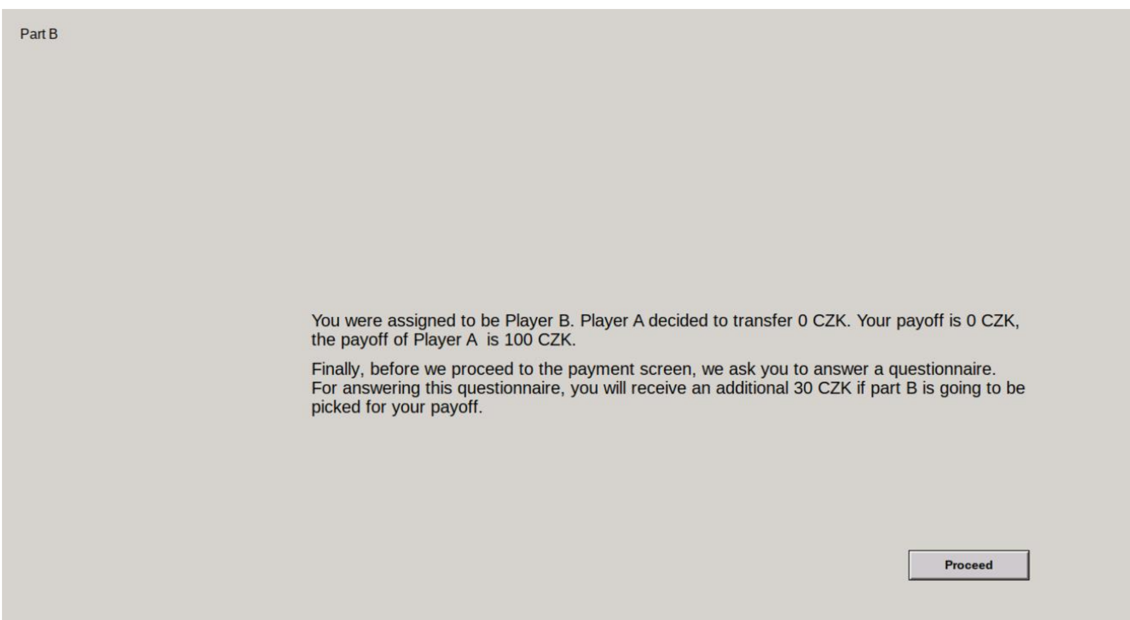
You will make a decision as Player A before you know if you are going to be assigned to be Player A or Player B.

Remember that there is a 20% chance that part B is going to be payoff relevant for you. If it is payoff relevant for you, your transfer as Player A will also be payoff relevant for Player B.

[screen 19 – part B]



[screen 20 – part B results]



[screen 21 – questionnaire]

Questionnaire
Please,fill in this short questionnaire:
Your answers do not influence your payoff.

Age

Sex: Male / Female

Nationality:

Study field:

Monthly net income in CZK

Highest degree earned: HighSchool Degree / Bachelor / Master / PhD

Have you participated in a similar game when one participant could decide on how much to transfer to another, who didn't get anything initially [Dictator Game]? Yes / No

Were you concentrated during the experiment? Not at all … Very much

Did you understand the instructions? Not at all … Very much

Why did/didn't you assign deduction points? (200 signs allowed)

Please, insert your unique code for payment here (you should obtain it in reminder email from hroot):

OK

[screen 22 – payment screen]

Results

The computer chose part A for the payment.

You guessed that other participants believe that 17.9 CZK should be transferred. The average response was 0.8 CZK.

For the second question, you guessed that on average, 35.1 CZK will be transferred. The average transfer was 23.8 CZK.

Therefore, you receive an additional 0 CZK.

Then, when assigning deduction point you received 50 CZK.

Player C transferred 0 CZK and you assigned 5 deduction points her/him.

After that, you guessed that other participants believe that 17.9 CZK should be transferred. The average response was 0.0 CZK.

For the second question, you guessed that 35.1 CZK was transferred. The average transfer was 23.8 CZK.

Therefore, you receive an additional 0 CZK.

In total, your payment from this experiment is 135 CZK (including base payment of 100 CZK).

Proceed

Instructions – punishees

[screen 1]

Experimental instructions

You are now taking part in an economic experiment. This experiment consists of one game that will be repeated for 4 rounds. You can earn a considerable amount of money depending on the decisions that you and the other participants make. Therefore, it is very important that you read the instructions carefully. It is important to us that you stay concentrated and in front of your computer. Communication with any of the participants is strictly forbidden and can lead to withholding of the payment.

You will receive your payoff on your bank account within two working days from the end of the experiment.

If you have questions or technical problems, please write to us through the chat in zoom.

[screen 2]

Experimental instructions

You will be paired randomly and anonymously with another participant. In each round, one of you will be randomly chosen to be Player C and the other Player D. Before making a decision, you will learn your role, which will be randomly assigned in each round anew.

Player C will receive 100 CZK and Player D will not receive anything. Player C can then decide to transfer either 0, 10, 40, or 50 CZK to Player D.

Other participants from this experiment (Players Y) have the opportunity to assign deduction points to Player C depending on Player C's transfer decisions. Each deduction point assigned to Player C will diminish Player C's income by 3 CZK. Players Y have to pay 1 CZK for each deduction point that they assign. They decide for each possible choice of transfer how many deduction points they want to assign to you. Before we start with the game, we will ask you about your emotions and opinions on the behavior of participants in this session.

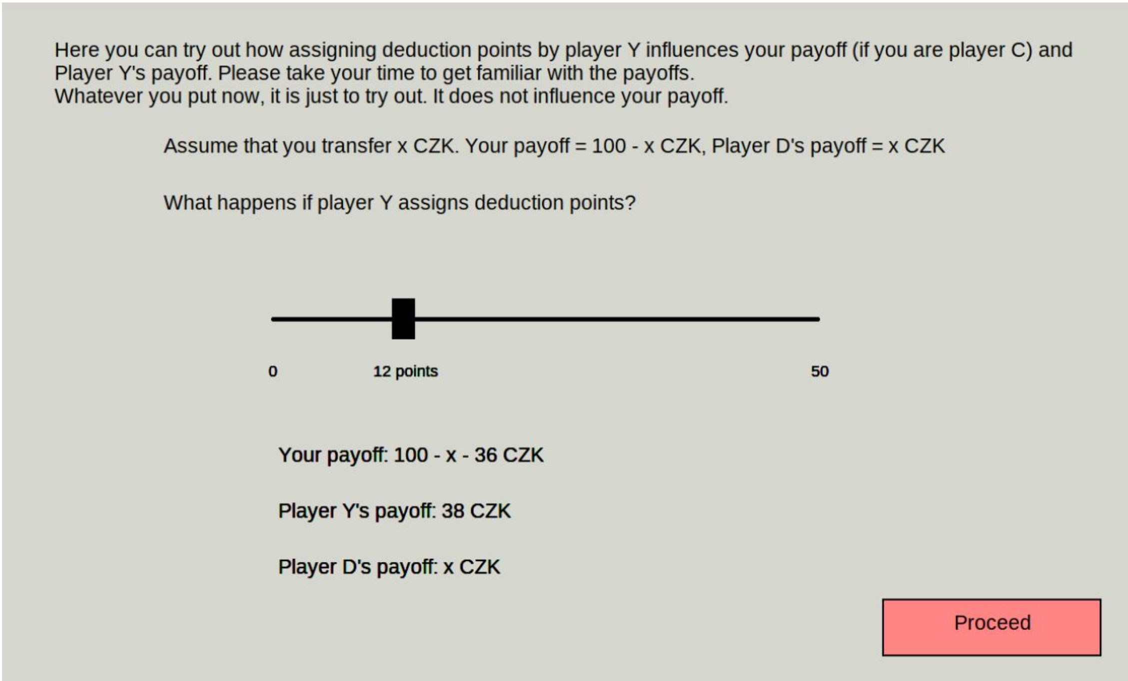
[screen 3 – tryout stage]

Here you can try out how assigning deduction points by player Y influences your payoff (if you are player C) and Player Y's payoff. Please take your time to get familiar with the payoffs.

Whatever you put now, it is just to try out. It does not influence your payoff.

Assume that you transfer x CZK. Your payoff = 100 - x CZK, Player D's payoff = x CZK

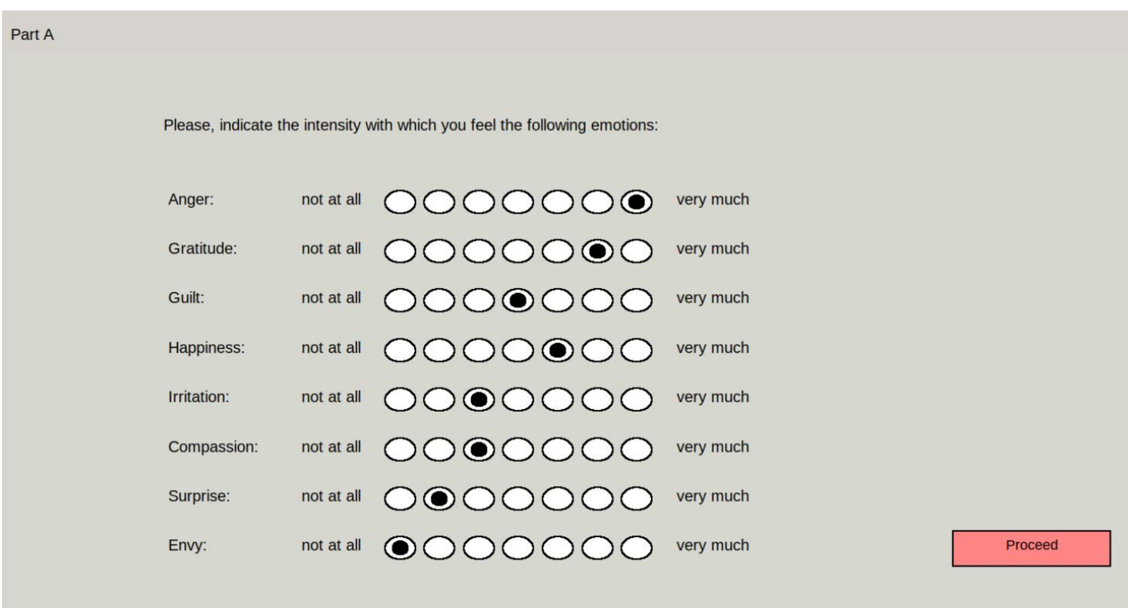
What happens if player Y assigns deduction points?



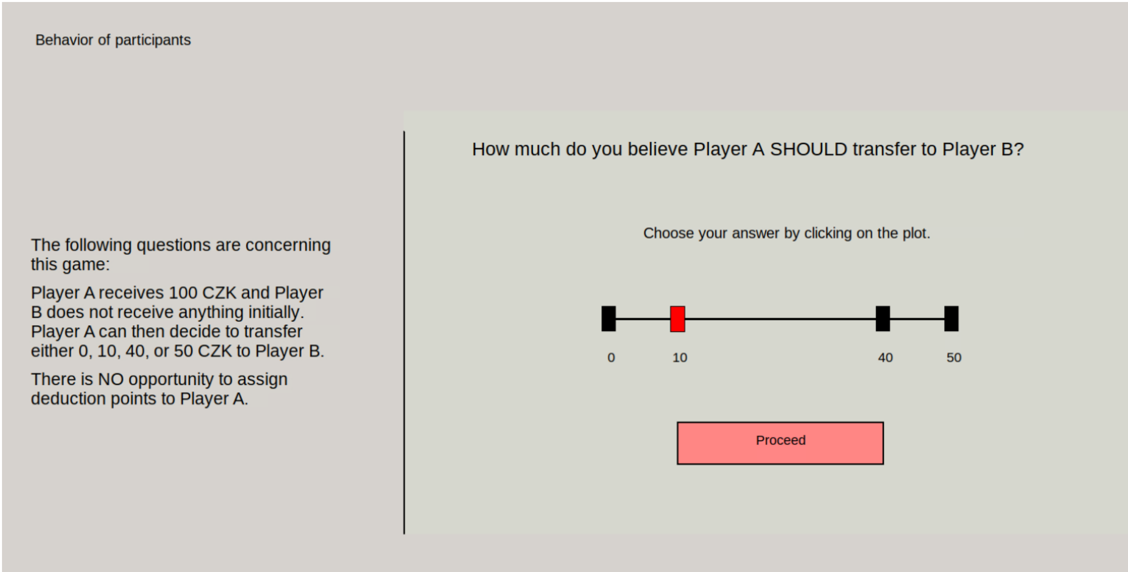
[screen 4]

Before we start with the experiment, we will ask you about your emotions and opinions on the behavior of participants in experimental session.

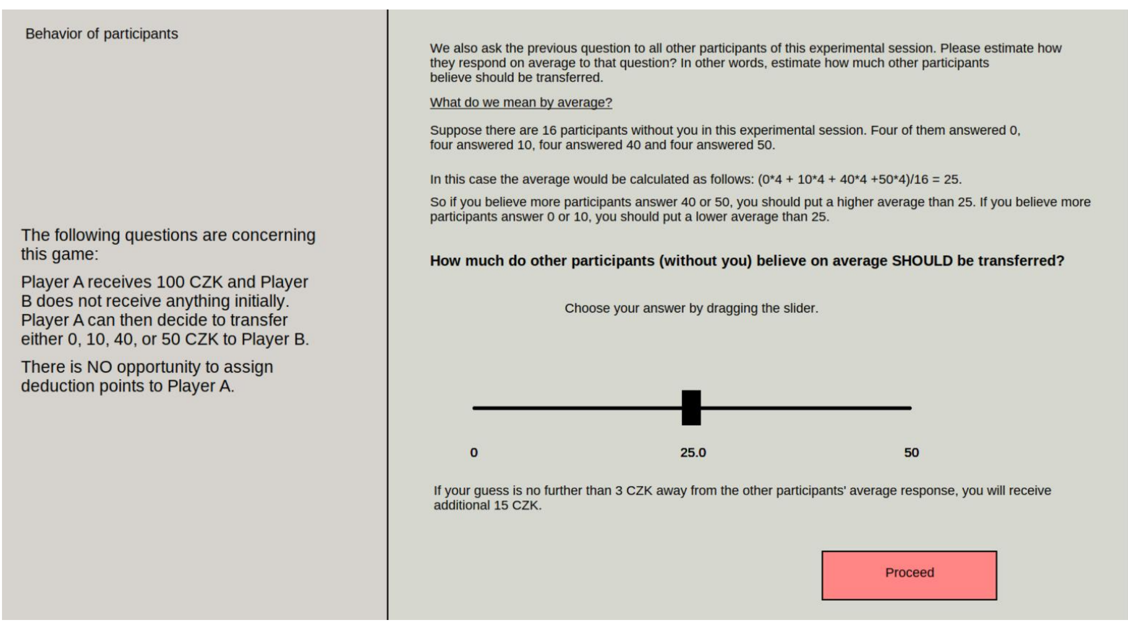
[screen 5 – emotions]



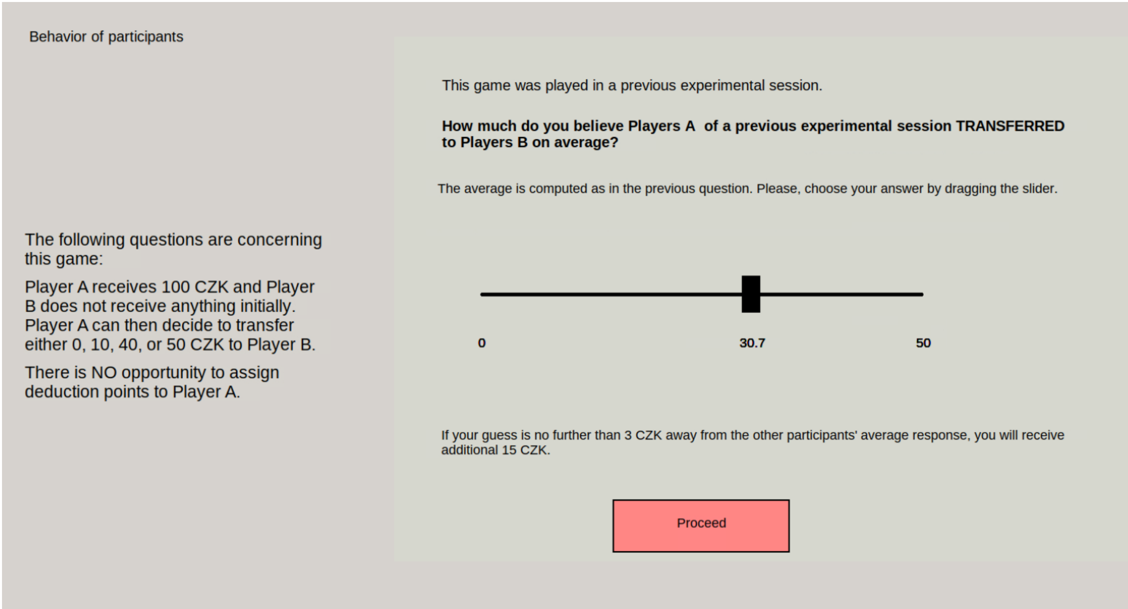
[screen 6 – norm elicitation – personal norm]



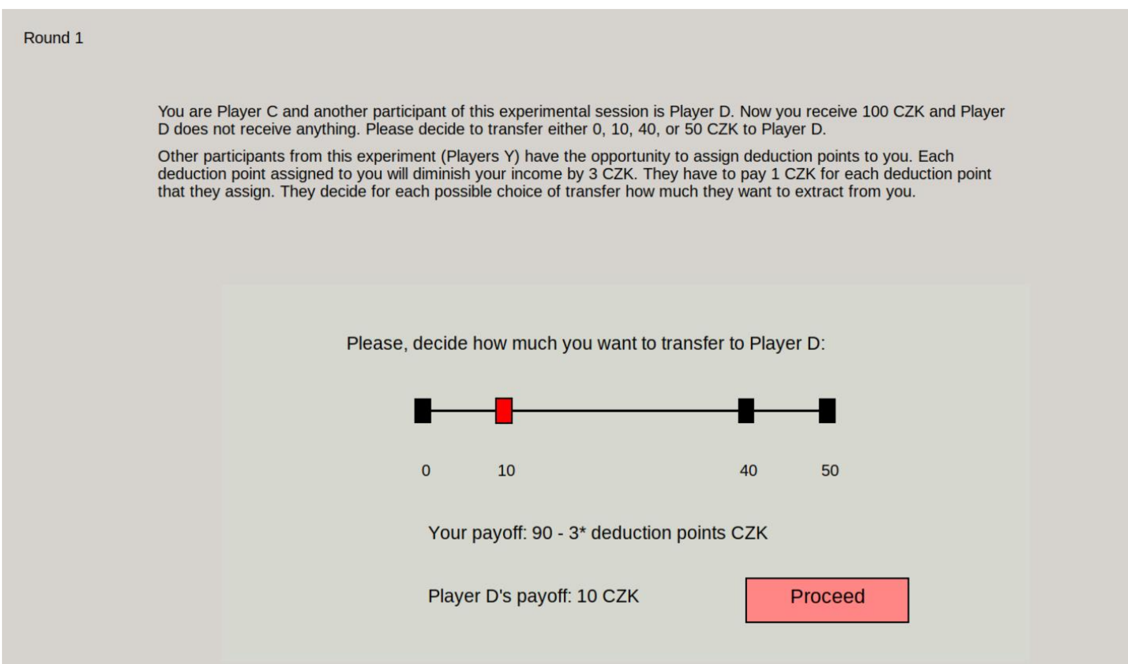
[screen 7 – norm elicitation – normative expectation]



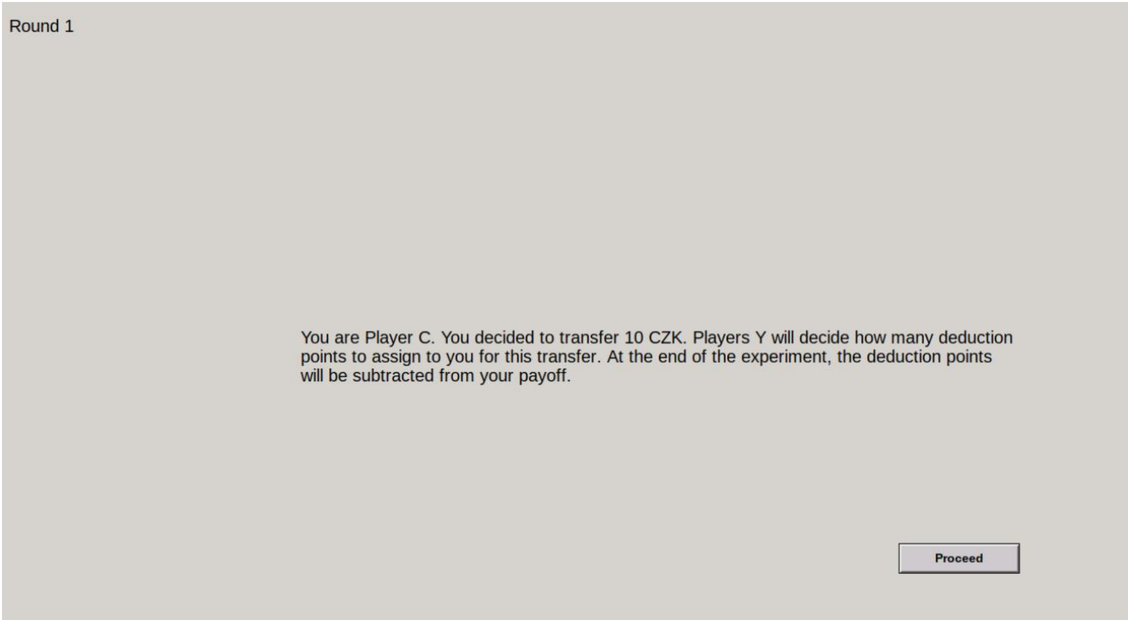
[screen 8 – norm elicitation – empirical expectation]



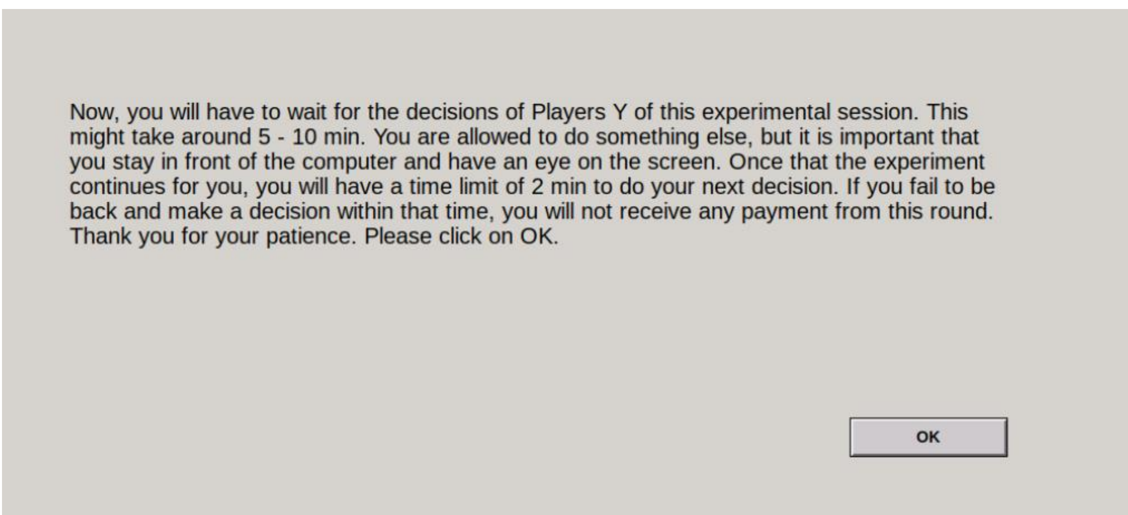
[screen 9 – dictator game round 1]



[screen 10 – dictator game results]

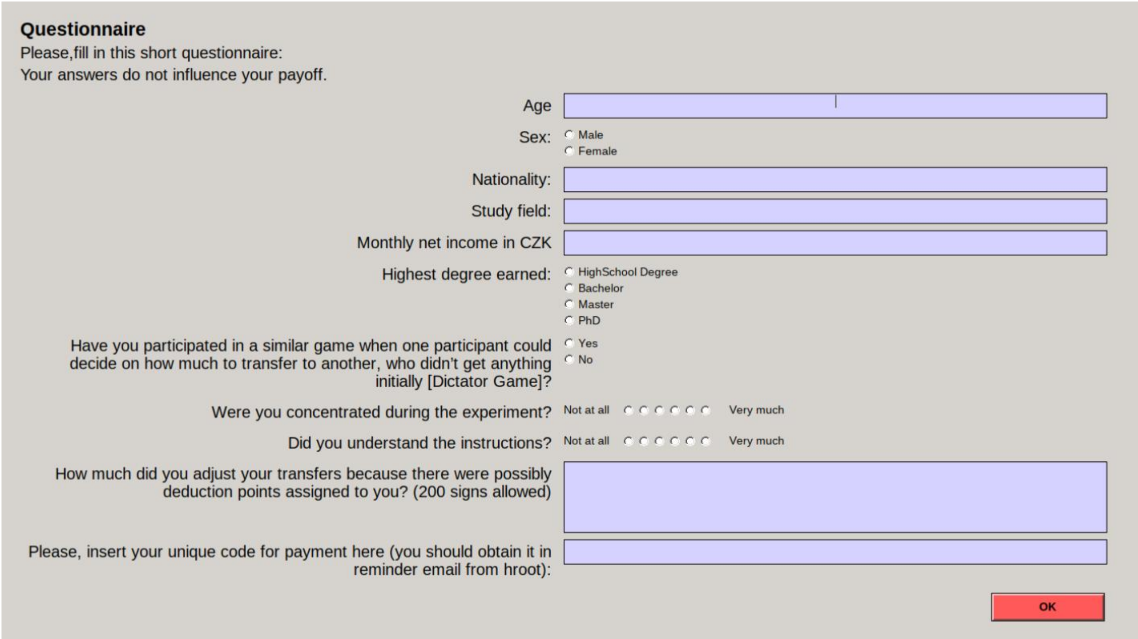


[screen 11 – waiting stage]



[screen 9 and 10 repeated 4 times]

[screen 11 – questionnaire]



Questionnaire

Please,fill in this short questionnaire:

Your answers do not influence your payoff.

Age

Sex: Male / Female

Nationality:

Study field:

Monthly net income in CZK

Highest degree earned: HighSchool Degree / Bachelor / Master / PhD

Have you participated in a similar game when one participant could decide on how much to transfer to another, who didn't get anything initially [Dictator Game]? Yes / No

Were you concentrated during the experiment? Not at all … Very much

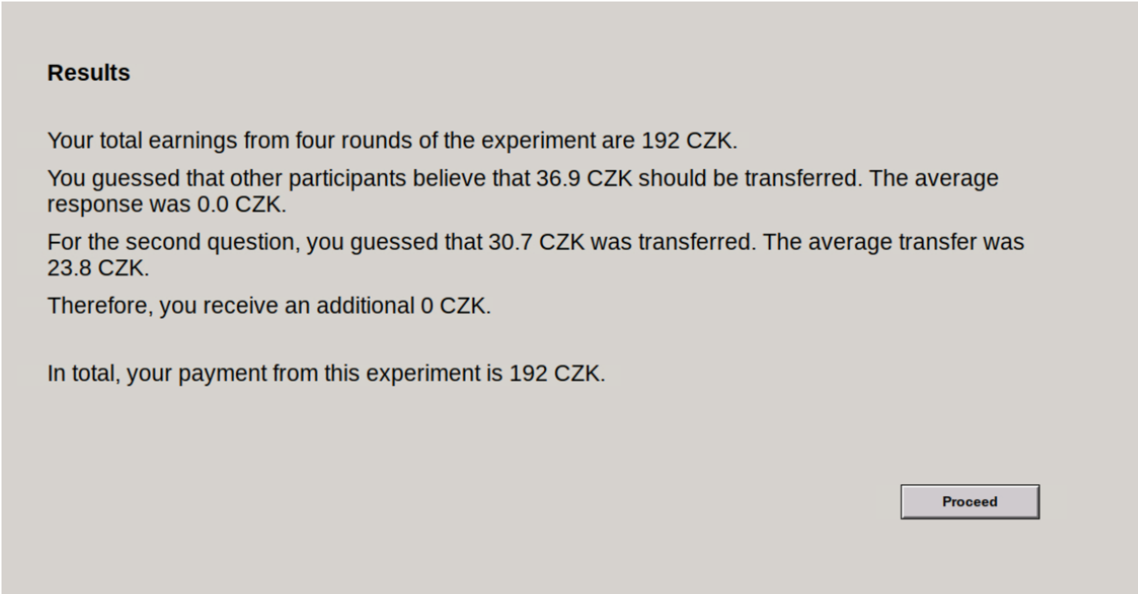
Did you understand the instructions? Not at all … Very much

How much did you adjust your transfers because there were possibly deduction points assigned to you? (200 signs allowed)

Please, insert your unique code for payment here (you should obtain it in reminder email from hroot):

OK

[screen 12 – payment screen]



Results

Your total earnings from four rounds of the experiment are 192 CZK.

You guessed that other participants believe that 36.9 CZK should be transferred. The average response was 0.0 CZK.

For the second question, you guessed that 30.7 CZK was transferred. The average transfer was 23.8 CZK.

Therefore, you receive an additional 0 CZK.

In total, your payment from this experiment is 192 CZK.

Proceed

Appendix D

Bibliography

Abbink, K. and Herrmann, B. (2011). The moral costs of nastiness. *Economic Inquiry*, 49(2):631–633.

Abbink, K. and Sadrieh, A. (2009). The pleasure of being nasty. *Economics Letters*, 105(3):306 – 308.

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Agerström, J., Carlsson, R., Nicklasson, L., and Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, 52:147–153.

Amegashie, J. A. (2012). Productive versus destructive efforts in contests. *European Journal of Political Economy*, 28(4):461–468.

Amegashie, J. A. et al. (2015). Sabotage in contests. *A companion to rent seeking and political economy*, pages 138–149.

Anbarci, N. and Feltovich, N. (2013). How sensitive are bargaining outcomes to changes in disagreement payoffs? *Experimental economics*, 16(4):560–596.

Anbarci, N. and Feltovich, N. (2018). How fully do people exploit their bargaining position? the effects of bargaining institution and the 50–50 norm. *Journal of Economic Behavior & Organization*, 145:320–334.

Anderson, D. R. (August 11, 1997). The cost of frivolous lawsuits is no joke. *Minneapolis St. Paul Business Journal*. Available at `https://www.bizjournals.com/twincities/stories/1997/08/11/editorial2.html`.

Anderson, L. R. and Stafford, S. L. (2003). An experimental analysis of rent seeking under varying competitive conditions. *Public Choice*, 115(1-2):199–216.

Andre, P., Boneva, T., Chopra, F., and Falk, A. (2024). Globally representative evidence on the actual and perceived support for climate action. *Nature Climate Change*, pages 1–7.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and ricardian equivalence. *Journal of Political Economy*, 97(6):1447–1458.

Andreoni, J. and Bernheim, B. D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J., Che, Y.-K., and Kim, J. (2007). Asymmetric information about rivals' types in standard auctions: An experiment. *Games and Economic Behavior*, 59(2):240–259.

Arechar, A. A., Gächter, S., and Molleman, L. (2018a). Conducting interactive experiments online. *Experimental Economics*, 21(1):99–131.

Arechar, A. A., Gächter, S., and Molleman, L. (2018b). Conducting interactive experiments online. *Experimental economics*, 21(1):99–131.

Aycinena, D. and Rentschler, L. (2019). Entry in contests with incomplete information: Theory and experiments. *European Journal of Political Economy*, 60:101803.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2018a). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503.

Azrieli, Y., Chambers, C. P., and Healy, P. J. (2018b). Incentives in experiments: A theoretical analysis. *Journal of Political Economy*, 126(4):1472–1503.

Balafoutas, L., Lindner, F., and Sutter, M. (2012). Sabotage in tournaments: Evidence from a natural experiment. *Kyklos*, 65(4):425–441.

Bartling, B., Gesche, T., and Netzer, N. (2017). Does the absence of human sellers bias bidding behavior in auction experiments? *Journal of the Economic Science Association*, 3(1):44–61.

Bašić, Z. and Verrina, E. (2023). Personal norms—and not only social norms—shape economic behavior. *MPI Collective Goods Discussion Paper*, (2020/25).

Bauer, M., Cahlíková, J., Celik Katreniak, D., Chytilová, J., Cingl, L., and Želinskỳ, T. (2023). Nastiness in groups. *Journal of the European Economic Association*, page jvad072.

Baumann, F. and Friehe, T. (2012). Emotions in litigation contests. *Economics of Governance*, 13(3):195–215.

Baye, M. R., Kovenock, D., and de Vries, C. G. (2005). Comparative analysis of litigation systems: An auction-theoretic approach. *The Economic Journal*, 115(505):583–601.

Baye, M. R., Kovenock, D., and de Vries, C. G. (2012). Contests with rank-order spillovers. *Economic Theory*, 51(2):315–350.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678.

Benistant, J. and Villeval, M. C. (2019). Unethical behavior and group identity in contests. *Journal of Economic Psychology*, 72:128–155.

Berinsky, A. J., Huber, G. A., and Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis*, 20(3):351.

Betto, M. and Thomas, M. W. (2024). Asymmetric all-pay auctions with spillovers. *Theoretical Economics*, 19(1):169–206.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C., Dimant, E., Gächter, S., and Nosenzo, D. (2022). Social proximity and the erosion of norm compliance. *Games and Economic Behavior*, 132:59–72.

Bicchieri, C., Dimant, E., Gelfand, M., and Sonderegger, S. (2023a). Social norms and behavior change: The interdisciplinary research frontier. *Journal of Economic Behavior Organization*, 205:A4–A7.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2023b). It's not a lie if you believe the norm does not apply: Conditional norm-following and belief distortion. *Games and Economic Behavior*, 138:321–354.

Bicchieri, C. and Xiao, E. (2009). Do the right thing: but only if others do so. *Journal of Behavioral Decision Making*, 22(2):191–208.

Bitkom (2018). Spionage, sabotage und datendiebstahl – wirtschaftsschutz in der industrie.

Boosey, L., Brookins, P., and Ryvkin, D. (2017). Contests with group size uncertainty: Experimental evidence. *Games and Economic Behavior*, 105:212–229.

Boosey, L., Brookins, P., and Ryvkin, D. (2020). Information disclosure in contests with endogenous entry: An experiment. *Management Science*, 66(11):5128–5150.

Bose, G., Dechter, E., and Ivancic, L. (2023). Conformity and adaptation in groups. *Journal of Economic Behavior & Organization*, 212:1267–1285.

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

Braeutigam, R., Owen, B., and Panzar, J. (1984). An economic analysis of alternative fee shifting systems. *Law and contemporary problems*, 47(2):173–185.

Brown, A. and Chowdhury, S. M. (2017). The hidden perils of affirmative action: Sabotage in handicap contests. *Journal of Economic Behavior & Organization*, 133:273–284.

Buhrmester, M., Kwang, T., and Gosling, S. D. (2011). Amazon's mechanical turk. *Perspectives on Psychological Science*, 6(1):3–5.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029.

Carbonara, E., Parisi, F., and von Wangenheim, G. (2015). Rent-seeking and litigation: The hidden virtues of limited fee shifting. *Review of Law & Economics*, 11(2):113–148.

Carpenter, J., Matthews, P. H., and Schirm, J. (2010). Tournaments and office politics: Evidence from a real effort experiment. *American Economic Review*, 100(1):504–17.

Carpenter, J. P. and Matthews, P. H. (2009). What norms trigger punishment? *Experimental Economics*, 12(3):272–288.

Carpenter, J. P. and Matthews, P. H. (2012). Norm enforcement: anger, indignation, or reciprocity? *Journal of the European Economic Association*, 10(3):555–572.

Carpenter, J. P., Matthews, P. H., and Ong'Ong'a, O. (2004). Why punish? social reciprocity and the enforcement of prosocial norms. *Journal of evolutionary economics*, 14:407–429.

Charness, G., Cobo-Reyes, R., and Jiménez, N. (2008). An investment game with third-party intervention. *Journal of Economic Behavior & Organization*, 68(1):18–28.

Charness, G., Masclet, D., and Villeval, M. C. (2014). The dark side of competition for status. *Management Science*, 60(1):38–55.

Chen, B., Galariotis, E., Ma, L., Wang, Z., and Zhu, Z. (2023). On disclosure of participation in innovation contests: a dominance result. *Annals of Operations Research*, 328:1615–1629.

Chen, B., Jiang, X., and Knyazev, D. (2017). On disclosure policies in all-pay auctions with stochastic entry. *Journal of Mathematical Economics*, 70:66–73.

Chen, B., Ma, L., Zhu, Z., and Zhou, Y. (2020a). Disclosure policies in all-pay auctions with bid caps and stochastic entry. *Economics Letters*, 186:108805.

Chen, B. and Rodrigues-Neto, J. A. (2023). The interaction of emotions and cost-shifting rules in civil litigation. *Economic Theory*, 75(3):841–885.

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

Chen, H., Zeng, Z., and Ma, J. (2020b). The source of punishment matters: Third-party punishment restrains observers from selfish behaviors better than does second-party punishment by shaping norm perceptions. *Plos one*, 15(3):e0229510.

Chen, K.-P. (2003). Sabotage in promotion tournaments. *Journal of Law, Economics, and Organization*, 19(1):119–140.

Chen, K.-P. and Wang, J.-S. (2007). Fee-shifting rules in litigation with contingency fees. *Journal of Law, Economics, & Organization*, 23(3):519–546.

Choi, A. and Sanchirico, C. (2004). Should plaintiffs win what defendants lose? litigation stakes, litigation effort, and the benefits of decoupling. *The Journal of Legal Studies*, 33(2):323–354.

Chowdhury, S., Danilov, A., and Kocher, M. (2022). Affirmative action policies and sabotage behavior in promotional tournaments: An experiment. Technical report, Working Paper.

Chowdhury, S. M., Esteve-González, P., and Mukherjee, A. (2023). Heterogeneity, leveling the playing field, and affirmative action in contests. *Southern Economic Journal*, 89(3):924–974.

Chowdhury, S. M. and Gürtler, O. (2015). Sabotage in contests: a survey. *Public Choice*, 164(1-2):135–155.

Chowdhury, S. M. and Sheremeta, R. M. (2011a). A generalized tullock contest. *Public Choice*, 147(3-4):413–420.

Chowdhury, S. M. and Sheremeta, R. M. (2011b). Multiple equilibria in tullock contests. *Economics Letters*, 112(2):216–219.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*, 58(6):1015.

Cooper, D. J. and Fang, H. (2008). Understanding overbidding in second price auctions: An experimental study. *The Economic Journal*, 118(532):1572–1595.

Cooper, D. J. and Kagel, J. (2016). *The handbook of experimental economics*, volume 2, chapter Other regarding preferences: A selective survey of experimental results, pages 217–290. Princeton University Press.

Coursey, D. L. and Stanley, L. R. (1988). Pretrial bargaining behavior within the shadow of the law: Theory and experimental evidence. *International Review of Law and Economics*, 8(2):161 – 179.

Crosetto, P. and Filippin, A. (2013). The "bomb" risk elicitation task. *Journal of risk and uncertainty*, 47(1):31–65.

Croson, R. T., Handy, F., and Shang, J. (2010). Gendered giving: the influence of social norms on the donation behavior of men and women. *International Journal of Nonprofit and Voluntary Sector Marketing*, 15(2):199–213.

Cubitt, R. P., Drouvelis, M., and Gächter, S. (2011). Framing and free riding: emotional responses and punishment in social dilemma games. *Experimental Economics*, 14(2):254–272.

d'Adda, G., Drouvelis, M., and Nosenzo, D. (2016). Norm elicitation in within-subject designs: Testing for order effects. *Journal of Behavioral and Experimental Economics*, 62:1–7.

Danilov, A. and Sliwka, D. (2017). Can contracts signal social norms? experimental evidence. *Management Science*, 63(2):459–476.

Dato, S. and Nieken, P. (2014). Gender differences in competition and sabotage. *Journal of Economic Behavior & Organization*, 100:64–80.

Dato, S. and Nieken, P. (2020). Gender differences in sabotage: the role of uncertainty and beliefs. *Experimental Economics*, 23(2):353–391.

Dechenaux, E., Kovenock, D., and Sheremeta, R. M. (2015). A survey of experimental research on contests, all-pay auctions and tournaments. *Experimental Economics*, 18(4):609–669.

Dechenaux, E. and Mancini, M. (2008). Auction-theoretic approach to modeling legal systems: An experimental analysis. *Applied Economics Research Bulletin*, 1:142–177.

Del Corral, J., Prieto-Rodriguez, J., and Simmons, R. (2010). The effect of incentives on sabotage: The case of spanish football. *Journal of Sports Economics*, 11(3):243–260.

DellaVigna, S. and Pope, D. (2018). What motivates effort? Evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069.

Deutscher, C., Frick, B., Gürtler, O., and Prinz, J. (2013). Sabotage in tournaments with heterogeneous contestants: Empirical evidence from the soccer pitch. *The Scandinavian Journal of Economics*, 115(4):1138–1157.

Dimant, E. and Gesche, T. (2023). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. *PNAS nexus*, 2(7):pgad224.

Duch, M. L., Grossmann, M. R., and Lauer, T. (2020). z-tree unleashed: A novel client-integrating architecture for conducting z-tree experiments over the internet. *Journal of Behavioral and Experimental Finance*, 28:100400.

Eisenkopf, G., Friehe, T., and Wohlschlegel, A. (2019). On the role of emotions in experimental litigation contests. *International Review of Law and Economics*, 57:90–94.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics*, 14:583–610.

Epley, N. and Gilovich, T. (2016). The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3):133–140.

Erkal, N., Gangadharan, L., and Koh, B. H. (2018). Monetary and non-monetary incentives in real-effort tournaments. *European Economic Review*, 101:528–545.

Evans, J. M., Hendron, M. G., and Oldroyd, J. B. (2015). Withholding the ace: The individual-and unit-level performance effects of self-reported and perceived knowledge hoarding. *Organization Science*, 26(2):494–510.

Fabbri, M. and Carbonara, E. (2017). Social influence on third-party punishment: An experiment. *Journal of Economic Psychology*, 62:204–230.

Faravelli, M., Friesen, L., and Gangadharan, L. (2015). Selection, tournaments, and dishonesty. *Journal of Economic Behavior & Organization*, 110:160–175.

Fehr, E. and Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, 25(2):63–87.

Feng, X. and Lu, J. (2016). The optimal disclosure policy in contests with stochastic entry: A bayesian persuasion perspective. *Economics Letters*, 147:103–107.

Fenn, P., Grembi, V., and Rickman, N. (2017). 'no win, no fee', cost-shifting and the costs of civil litigation: A natural experiment. *The Economic Journal*, 127(605):F142–F163.

Filiz-Ozbay, E. and Ozbay, E. Y. (2007). Auctions with anticipated regret: Theory and experiment. *American Economic Review*, 97(4):1407–1418.

Filiz-Ozbay, E. and Ozbay, E. Y. (2010). Anticipated loser regret in third price auctions. *Economics Letters*, 107(2):217–219.

Fišar, M., Kubák, M., Špalek, J., and Tremewan, J. (2016). Gender differences in beliefs and actions in a framed corruption experiment. *Journal of Behavioral and Experimental Economics*, 63:69–82.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Fischer, S., Güth, W., and Pull, K. (2007). Is there as-if bargaining? *The Journal of Socio-Economics*, 36(4):546–560.

Ford, D. P. and Staples, S. (2010). Are full and partial knowledge sharing the same? *Journal of knowledge management*, 14(3):394–409.

Fu, Q., Jiao, Q., and Lu, J. (2011). On disclosure policy in contests with stochastic entry. *Public Choice*, 148(3-4):419–434.

Fu, Q., Jiao, Q., and Lu, J. (2015). Contests with endogenous entry. *International Journal of Game Theory*, 44(2):387–424.

Fu, Q., Lu, J., and Zhang, J. (2016). Disclosure policy in tullock contests with asymmetric stochastic entry. *Canadian Journal of Economics/Revue canadienne d'économique*, 49(1):52–75.

Gabuthy, Y., Peterle, E., and Tisserand, J.-C. (2021). Legal fees, cost-shifting rules and litigation: experimental evidence. *Journal of Behavioral and Experimental Economics*, 93:101705.

Gino, F., Ayal, S., and Ariely, D. (2009). Contagion and differentiation in unethical behavior: The effect of one bad apple on the barrel. *Psychological science*, 20(3):393–398.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Gu, Y., Hehenkamp, B., and Leininger, W. (2019). Evolutionary equilibrium in contests with stochastic participation: Entry, effort and overdissipation. *Journal of Economic Behavior & Organization*, 164:469–485.

Guha, B. (2016). Malicious litigation. *International Review of Law and Economics*, 47(C):24–32.

Guha, B. (2018). Malice in the rubinstein bargaining game. *Mathematical Social Sciences*, 94:82–86.

Guha, B. (2019). Malice in pretrial negotiations. *International Review of Law and Economics*, 58:25–33.

Gürtler, O. and Münster, J. (2010). Sabotage in dynamic tournaments. *Journal of Mathematical Economics*, 46(2):179–190.

Gürtler, O., Münster, J., and Nieken, P. (2013). Information policy in tournaments with sabotage. *The Scandinavian Journal of Economics*, 115(3):932–966.

Harbring, C. and Irlenbusch, B. (2005). Incentives in tournaments with endogenous prize selection. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pages 636–663.

Harbring, C. and Irlenbusch, B. (2008). How many winners are good to have?: On tournaments with sabotage. *Journal of Economic Behavior & Organization*, 65(3-4):682–702.

Harbring, C. and Irlenbusch, B. (2011). Sabotage in tournaments: Evidence from a laboratory experiment. *Management Science*, 57(4):611–627.

Harbring, C., Irlenbusch, B., Kräkel, M., and Selten, R. (2007). Sabotage in corporate contests–an experimental analysis. *International Journal of the Economics of Business*, 14(3):367–392.

Hause, J. C. (1989). Indemnity, settlement, and litigation, or i'll be suing you. *The Journal of Legal Studies*, 18(1):157–79.

Helland, E. and Yoon, J. (2017). Estimating the effects of the english rule on litigation outcomes. *Review of Economics and Statistics*, 99(4):678–682.

Helmers, C., Lefouili, Y., Love, B. J., and McDonagh, L. (2021). The effect of fee shifting on litigation: evidence from a policy innovation in intermediate cost shifting. *American Law and Economics Review*, 23(1):56–99.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., et al. (2006). Costly punishment across human societies. *Science*, 312(5781):1767–1770.

Herrmann, B. and Orzen, H. (2008). The appearance of homo rivalis: Social preferences and the nature of rent seeking. Technical report, CeDEx discussion paper series.

Heyes, A., Rickman, N., and Tzavara, D. (2004). Legal expenses insurance, risk aversion and litigation. *International Review of Law and Economics*, 24(1):107 – 119.

Higgins, R. S., Shughart, W. F., and Tollison, R. D. (1988). *Free Entry and Efficient Rent-Seeking*, pages 127–139. Springer US, Boston, MA.

Hirshleifer, J. and Osborne, E. (2001). Truth, effort, and the legal battle. *Public Choice*, 108(1):169–195.

Hoeft, L., Mill, W., and Vostroknutov, A. (2023). Normative perception of power abuse. *MPI Collective Goods Discussion Paper*, (2019/6).

Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, 92(5):1644–1655.

Horton, J. J., Rand, D. G., and Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, 14(3):399 – 425.

House, B. R., Kanngiesser, P., Barrett, H. C., Yilmaz, S., Smith, A. M., Sebastian-Enesco, C., Erut, A., and Silk, J. B. (2020). Social norms and cultural diversity in the development of third-party punishment. *Proceedings of the Royal Society B*, 287(1925):20192794.

Hughes, J. W. and Snyder, E. A. (1995). Litigation and settlement under the english and american rules: Theory and evidence. *The Journal of Law & Economics*, 38(1):225–250.

Inglis, L., McCabe, K., Rassenti, S., and Simmons, D. (2005). Experiments on the effects of cost-shifting, court costs, and discovery on the efficient settlement of tort claims. *Florida State University Law Review*, 33:89.

Jiao, Q., Ke, C., and Liu, Y. (2022). When to disclose the number of contestants: Theory and experimental evidence. *Journal of Economic Behavior & Organization*, 193:146–160.

Jordan, J., McAuliffe, K., and Rand, D. (2016a). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, 19(4):741–763.

Jordan, J. J., Hoffman, M., Bloom, P., and Rand, D. G. (2016b). Third-party punishment as a costly signal of trustworthiness. *Nature*, 530(7591):473–476.

Jordan, J. J., Sommers, R., Bloom, P., and Rand, D. G. (2017). Why do we hate hypocrites? Evidence for a theory of false signaling. *Psychological Science*, 3(1):356–368.

Kahana, N. and Klunover, D. (2015). A note on poisson contests. *Public Choice*, 165(1-2):97–102.

Kahana, N. and Klunover, D. (2016). Complete rent dissipation when the number of rent seekers is uncertain. *Economics Letters*, 141:8–10.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the coase theorem. *Journal of Political Economy*, 98(6):1325–1348.

Kahneman, D. and Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, 39.

Kamei, K. (2018). The role of visibility on third party punishment actions for the enforcement of social norms. *Economics letters*, 171:193–197.

Kamei, K. (2020). Group size effect and over-punishment in the case of third party enforcement of social norms. *Journal of Economic Behavior & Organization*, 175:395–412.

Kamei, K., Sharma, S., and Walker, M. J. (2023). Sanction enforcement among third parties: New experimental evidence from two societies. *Available at SSRN 4429802*.

Katz, A. W. and Sanchirico, C. W. (2010). Fee shifting in litigation: survey and assessment. *University of Pennsylvania Institute for Law and Economics Research Paper*, (10-30).

Keizer, K., Lindenberg, S., and Steg, L. (2008). The spreading of disorder. *Science*, 322(5908):1681–1685.

Kessler, J. B. and Leider, S. (2012). Norms and contracting. *Management Science*, 58(1):62–77.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences*. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Kimbrough, E. O. and Reiss, J. P. (2012). Measuring the distribution of spitefulness. *PLOS ONE*, 7(8):1–8.

Kirchkamp, O. and Mill, W. (2021a). Spite vs. risk: Explaining overbidding in the second-price all-pay auction. *Games and Economic Behavior*, 130:616–635.

Kirchkamp, O. and Mill, W. (2021b). Spite vs. risk: Explaining overbidding in the second-price all-pay auction: A theoretical and experimental investigation. *Games and Economic Behavior*, 130:616–635.

Kisner, P. (1976). Malicious prosecution: An effective attack on spurious medical malpractice claims. *Case Western Reserve Law Review*, 26(3):653–686.

Kölle, F. and Quercia, S. (2021). The influence of empirical and normative expectations on cooperation. *Journal of Economic Behavior & Organization*, 190:691–703.

Konrad, K. A. (2000). Sabotage in rent-seeking contests. *Journal of Law, Economics, and Organization*, 16(1):155–165.

Kräkel, M. (2005). Helping and sabotaging in tournaments. *International Game Theory Review*, 7(02):211–228.

Kromer, E. and Bahçekapili, H. G. (2010). The influence of cooperative environment and gender on economic decisions in a third party punishment game. *Procedia-Social and Behavioral Sciences*, 5:250–254.

Krysowski, E. and Tremewan, J. (2021). Why does anonymity make us misbehave: Different norms or less compliance? *Economic Inquiry*, 59(2):776–789.

Kumar Jha, J. and Varkkey, B. (2018). Are you a cistern or a channel? exploring factors triggering knowledge-hiding behavior at the workplace: evidence from the indian r&d professionals. *Journal of Knowledge Management*, 22(4):824–849.

Lau, R. R. and Rovner, I. B. (2009). Negative campaigning. *Annual review of political science*, 12:285–306.

Lazear, E. P. (1989). Pay equality and industrial politics. *Journal of Political Economy*, 97(3):561–580.

Legal Information Institute (2018). Vexatious litigation. *Legal Information Institute*. Available at `https://www.law.cornell.edu/wex/vexatious_litigation`.

Leibbrandt, A. and López-Pérez, R. (2012). An exploration of third and second party punishment in ten simple games. *Journal of Economic Behavior & Organization*, 84(3):753–766.

Leibbrandt, A., Wang, L. C., and Foo, C. (2017). Gender quotas, competitions, and peer review: Experimental evidence on the backlash against women. *Management Science*, 64(8):3501–3516.

Lergetporer, P., Angerer, S., Glätzle-Rützler, D., and Sutter, M. (2014). Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation. *Proceedings of the National Academy of Sciences*, 111(19):6916–6921.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, 1(3):593–622.

Li, X., Molleman, L., and van Dolder, D. (2021). Do descriptive social norms drive peer punishment? conditional punishment strategies and their impact on cooperation. *Evolution and Human Behavior*, 42(5):469–479.

Lim, W. and Matros, A. (2009). Contests with a stochastic number of players. *Games and Economic Behavior*, 67(2):584–597.

Lois, G. and Wessa, M. (2019). Creating sanctioning norms in the lab: The influence of descriptive norms in third-party punishment. *Social Influence*, 14(2):50–63.

Main, B. and Park, A. (2000). The british and american rules: an experimental examination of pre-trial bargaining within the shadow of the law. *Scottish Journal of Political Economy*, 47:37–60.

Main, B. G. and Park, A. (2002). The impact of defendant offers into court on negotiation in the shadow of the law: experimental evidence. *International Review of Law and Economics*, 22(2):177 – 192.

Mao, A., Dworkin, L., Suri, S., and Watts, D. J. (2017). Resilient cooperators stabilize long-run cooperation in the finitely repeated prisoner's dilemma. *Nature Communications*, 8(13800).

March, C. and Sahm, M. (2017). Asymmetric discouragement in asymmetric contests. *Economics Letters*, 151:23–27.

Marcus, D. K., Zeigler-Hill, V., Mercer, S. H., and Norris, A. L. (2014). The psychology of spite and the measurement of spitefulness. *Psychological Assessment*, 26(2):563–574.

Martin, J. W., Martin, S., and McAuliffe, K. (2021). Third-party punishment promotes fairness in children. *Developmental Psychology*, 57(6):927.

Massenot, B., Maraki, M., and Thöni, C. (2021). Litigation spending and care under the english and american rules: Experimental evidence. *American Law and Economics Review*, 23(1):164–206.

Mathew, S. and Boyd, R. (2011). Punishment sustains large-scale cooperation in prestate warfare. *Proceedings of the National Academy of Sciences*, 108(28):11375–11380.

McAuliffe, K., Jordan, J. J., and Warneken, F. (2015). Costly third-party punishment in young children. *Cognition*, 134:1–10.

McBride, M. and Skaperdas, S. (2014). Conflict, settlement, and the shadow of the future. *Journal of Economic Behavior & Organization*, 105(C):75–89.

McBride, M., Skaperdas, S., and Tsai, P.-H. (2017). Why go to court? bargaining failure under the shadow of trial with complete information. *European Journal of Political Economy*.

Merguei, N., Strobel, M., and Vostroknutov, A. (2022). Moral opportunism as a consequence of decision making under uncertainty. *Journal of Economic Behavior & Organization*, 197:624–642.

Mill, W. (2017a). The spite motive in third price auctions. *Economics Letters*, 161:71–73.

Mill, W. (2017b). The spite motive in third price auctions. *Economics Letters*, 161:71 – 73.

Mill, W. and Morgan, J. (2018). Competition between friends and foes. Mimeo.

Mill, W. and Morgan, J. (2021). The cost of a divided america: an experimental study into destructive behavior. *Experimental Economics*.

Mill, W. and Morgan, J. (2022a). Competition between friends and foes. *European Economic Review*, 147:104171.

Mill, W. and Morgan, J. (2022b). The cost of a divided america: an experimental study into destructive behavior. *Experimental Economics*, 25:974–1001.

Mill, W. and Stäbler, J. (2023). Spite in litigation. *CESifo Working Paper No. 10290,*.

Montero, M. (2008). Altruism, spite and competition in bargaining games. *Theory and Decision*, 65(2):125–151.

Morgan, J., Orzen, H., and Sefton, M. (2012). Endogenous entry in contests. *Economic Theory*, 51(2):435–463.

Morgan, J., Steiglitz, K., and Reis, G. (2003a). The spite motive and equilibrium behavior in auctions. *Contributions in Economic Analysis & Policy*, 2(1):1–25.

Morgan, J., Steiglitz, K., and Reis, G. (2003b). The Spite Motive and Equilibrium Behavior in Auctions. *Contributions in Economic Analysis & Policy*, 2(1).

Münster, J. (2006). Contests with an unknown number of contestants. *Public Choice*, 129(3-4):353–368.

Münster, J. (2007). Selection tournaments, sabotage, and participation. *Journal of Economics & Management Strategy*, 16(4):943–970.

Murphy, R. O. and Ackerman, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1):13–41.

Murphy, R. O., Ackerman, K. A., and Handgraaf, M. J. J. (2011a). Measuring social value orientation. *Judgment and Decision Making*, 6(8):771–781.

Murphy, R. O. and Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1):13–41.

Murphy, R. O., Ackermann, K. A., and Handgraaf, M. (2011b). Measuring social value orientation. *Judgment and Decision making*, 6(8):771–781.

Myerson, R. B. and Wärneryd, K. (2006). Population uncertainty in contests. *Economic Theory*, 27(2):469–474.

Nash, J. F. (1950). The bargaining problem. *Econometrica*, 18(2):155–162.

Nelissen, R. M. and Zeelenberg, M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt and the functions of altruistic sanctions. *Judgment and Decision making*, 4(7):543.

Nissen, M. and Haugsted, F. (2020). Badmouthing your competitor's products: When does denigration become an antitrust issue?

Pan, W., Zhang, Q., Teo, T. S., and Lim, V. K. (2018). The dark triad and knowledge hiding. *International Journal of Information Management*, 42:36–48.

Paolacci, G., Chandler, J., and Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making*, 5(5):411–419.

Parisi, F. (2002). Rent-seeking through litigation: adversarial and inquisitorial systems compared. *International Review of Law and Economics*, 22(2):193 – 216.

Peysakhovich, A., Nowak, M. A., and Rand, D. G. (2014). Humans display a 'cooperative phenotype' that is domain general and temporally stable. *Nature Communications*, 5:4939.

Philippi, M. J. (1983). Malicious prosecution and medical malpractice legislation in indiana: A quest for balance. *Valparaiso University Law Review*, 17(4):877–909.

Piardini, P., Drouvelis, M., and Di Cagno, D. (2017). Gender effects and third-party punishment in social dilemma games.

Piest, S. and Schreck, P. (2021). Contests and unethical behavior in organizations: a review and synthesis of the empirical literature. *Management Review Quarterly*, 71(4):679–721.

Plott, C. R. (1987). Legal fees: A comparison of the american and english rules. *Journal of Law, Economics, & Organization*, 3(2):185–192.

Plott, C. R. and Zeiler, K. (2007). Exchange asymmetries incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review*, 97(4):1449–1466.

Post, A. (July 22, 2011). Frivolous lawsuits clogging u.s. courts, stalling economic growth. *Law.com*. Available at `https://www.law.com/almID/4de52195140ba07e6e000573/?slreturn=20180203090120`.

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., and Greene, J. D. (2014). Social heuristics shape intuitive cooperation. *Nature Communications*, 5:3677.

Reinganum, J. F. and Wilde, L. L. (1986). Settlement, litigation, and the allocation of litigation costs. *The RAND Journal of Economics*, 17(4):557–566.

Reuben, E. and Riedl, A. (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior*, 77(1):122–137.

Ryvkin, D. and Drugov, M. (2020). The shape of luck and competition in winner-take-all tournaments. *Theoretical Economics*, 15(4):1587–1626.

Schmidt, R. J. (2019). Do injunctive or descriptive social norms elicited using coordination games better explain social preferences? Technical report, Discussion Paper Series.

Schweizer, U. (1989). Litigation and settlement under two-sided incomplete information. *The Review of Economic Studies*, 56(2):163–177.

Selten, R. (1967). *Beiträge zur experimentellen Wirtschaftforschung*, chapter Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes, pages 136–168. Mohr, Tübingen.

Serenko, A. (2020). Knowledge sabotage as an extreme form of counterproductive knowledge behavior: the perspective of the target. *Journal of Knowledge Management*, 24(4):737–773.

Sheremeta, R. M. (2011). Contest design: An experimental investigation. *Economic Inquiry*, 49(2):573–590.

Sheremeta, R. M. (2013). Overbidding and heterogeneous behavior in contest experiments. *Journal of Economic Surveys*, 27(3):491–514.

Sheremeta, R. M. (2018). Impulsive behavior in competition: Testing theories of overbidding in rent-seeking contests. *Available at SSRN 2676419*.

Skaperdas, S. (1996). Contest success functions. *Economic theory*, 7(2):283–290.

Snyder, E. A. and Hughes, J. W. (1990). The english rule for allocating legal costs: Evidence confronts theory. *Journal of Law, Economics, & Organization*, 6(2):345–380.

Spier, K. E. (1992). The dynamics of pretrial negotiation. *The Review of Economic Studies*, 59(1):93–108.

Spier, K. E. (1994). Pretrial bargaining and the design of fee-shifting rules. *The RAND Journal of Economics*, 25(2):197–214.

Spier, K. E. (2007). Litigation. In Polinsky, M. and Shavell, S., editors, *Handbook of Law and Economics*, volume 1, chapter 4, pages 259 – 342. North Holland, North Holland, 1 edition.

Suri, S. and Watts, D. J. (2011). Cooperation and contagion in web-based, networked public goods experiments. *PLOS ONE*, 6(3):1–18.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., and Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *American Economic Review*, 103(1):510–31.

Thomas, R. E. (1995). The trial selection hypothesis without the 50 percent rule: Some experimental evidence. *The Journal of Legal Studies*, 24(1):209–228.

Tremewan, J. and Vostroknutov, A. (2021). An informational framework for studying social norms. In *A research agenda for experimental economics*, pages 19–42. Edward Elgar Publishing.

Tullock, G. (1980). Efficient rent seeking. In James M. Buchanan, Robert D. Tollison, . G. T., editor, *Toward a theory of rent-seeking society*, pages 97–112. College Station: Texas AM University Press.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Vandegrift, D. and Yavas, A. (2010). An experimental test of sabotage in tournaments. *Journal of Institutional and Theoretical Economics JITE*, 166(2):259–285.

Wang, X. and Liu, S. (2023). Disclosure policies in all-pay auctions with bid caps and stochastic entry: The impact of risk aversion. *Bulletin of Economic Research*.

Yago, G. (January, 1999). The economic costs of frivolous securities litigation. *Milken Institute Review*.

Zimmermann, F. (2020). The dynamics of motivated beliefs. *American Economic Review*, 110(2):337–363.

Zong, J., De Jong, E., Qiu, J., and Li, J. (2021). Socially appropriate intervention: A cross-country investigation of third-party norm enforcement. *Available at SSRN 3943578*.

Eidesstattliche Versicherung gemäß §8 Absatz 2 Buchstabe a) der Promotionsordnung der Universität Mannheim zur Erlangung des Doktorgrades der Volkswirtschaftslehre (Dr. rer. pol.)

1. Ich habe die vorliegende Arbeit selbständig angefertigt und die benutzten Hilfsmittel vollständig und deutlich angegeben.

2. Die eingereichten Dissertationsexemplare sowie der Datenträger gehen in das Eigentum der Universität über.

Mannheim, 13.05.2024

Jonathan Stäbler

Curriculum Vitae

2018 - 2024	University of Mannheim Center for Doctoral Studies in Economics Ph.D. in Economics (Dr. rer. pol.)
2015 - 2017	Karlsruhe Instistute of Technology Master of Science in Industrial Engineering and Management
Fall 2016	University of Massachusetts Amherst Exchange Student
2011 - 2015	Karlsruhe Institute of Technology Bachelor of Science in Industrial Engineering and Management
Fall 2013	Universitat de Barcelona Exchange Student