Behavioral Economic Essays on Image Concerns and Social Norms in Moral Decision-Making and Negotiations

Universität Hamburg

Fakultät für Wirtschafts- und Sozialwissenschaften

Kumulative Dissertation

Zur Erlangung der Würde des Doktors der

Wirtschafts- und Sozialwissenschaften

„Dr. rer. pol.“

(gemäß der PromO vom 18. Januar 2017)

vorgelegt von

Litsios, Christos

aus Meerbusch, Deutschland

Hamburg, den 11. Februar 2026

Vorsitzende/r: Prof. Dr. Andreas Lange

Erstgutachter/in: Prof. Dr. Gerd Mühlheußer

Zweitgutachter/in: Prof. Dr. Dr. Lydia Mechtenberg

Datum der Disputation: 11.06.2026

To the vividly unfinished

Acknowledgements

Considering one of the many great lessons this dissertation has so generously taught me in life, a project like this might be executed by one person, but not without the great support of many caring others.

First, I would like to thank my supervisors, Gerd Mühlheußer and Lydia Mechtenberg, and my panel chair, Andreas Lange. All my economic studies, from my Bachelor's degree to this PhD thesis, were completed at the University of Hamburg. I carry vivid memories of my first encounters as a student attending your lectures. All three of you have deeply inspired me and, in doing so, significantly contributed to the emergence of a dream that has become reality in this thesis. I thank you all for sharing your passion with me. I would like to thank you, Gerd, for this long journey under your generous and sensitive supervision—from first initiating me into the world of academia, through highs and lows, to sharing your understanding and motivating me to follow through to the end. Thank you for supporting me in many ways, especially for encouraging me to follow my own ideas and thereby helping me develop confidence in my work. Lydia, thank you for your thoughtful feedback and for setting an example by embracing independent and fearless thinking. Andreas, thank you for your great support as panel chair.

Second, I would like to thank my co-authors Fanny Schories, Timo Promann, Annika Nieper, and Christoph Huber, as well as my colleagues Aljoscha Minnich, Jan Biermann, and Bastian Westbrook, for their contributions, for teaching me, and for growing with me along this path. I would like to thank you, Fanny, for being bold and pursuing our genuine ideas, for carefully examining and engaging with the storm of my thoughts, for teaching me rigorous pragmatism, and for your patience. Thank you, Timo, Annika, and Christoph, for being an amazing team and for generously sharing your knowledge, experience, and ideas. Thank you, Aljoscha, not only for being a bright and insightful yet humble colleague with an open ear and just the right questions at the right moment, but also for being a friend with whom to surf the tides of life's ups and downs. Jan, thank you for being a friend and for teaching me—among many other things—to genuinely think outside the box, with your inquisitive mind pointing out possibilities beyond my scope. I would like to thank you, Bastian, for being there at the right moment, for making time to listen with an open and curious mind, and for your careful feedback. Your willingness to make your own process of weighing theoretical thoughts transparent was tremendously important in helping me learn to scrutinize my own ideas.

I would also like to thank the University of Hamburg for being a place where my curiosity was met with so many possibilities and inspiring people. I thank the WiSo Graduate School for its support and courses, as well as the WiSo-Experimentallabor, where I spent many hours and was always supported by a great team and leadership. A number of institutions further supported my research financially through grants, summer schools, and conferences, including the German Association for Experimental Economic Research (GfeW), the German Research Foundation (DFG), the Foundation of the Swiss National Bank, the Summer School in Behavioral Game Theory 2020 by David J. Cooper, the Summer School on Psychological Game Theory, the Young Economists' Meeting 2022 in Brno, and the British University in Egypt.

Finally, I would like to use this moment to thank the many people outside of academia who supported me throughout this endeavor. In particular, I would like to thank Felix, Maximilian, Anna, Natascha, and Torsten for their presence, compassionate listening, and understanding. I thank the Contact Improvisation community, where I learned not only to accept but to embrace failing. And lastly, thank you, Jule, for being my loving and understanding partner, who supports me and accepts me for who I am, with all that comes along.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Behavioral economics has developed since the mid-20th century into a field of economics that integrates psychological insights into the analysis of economic decision-making. Although psychological thought has been at the heart of economics from the very beginning, as writings by Adam Smith, Jeremy Bentham, and Francis Edgeworth show, there has been somewhat of a break around the turn of the 20th century. With the neoclassical revolution, economists hoped for the discipline to become more of a natural science. This led to a split with the less developed, scientific discipline of psychology at the time Camerer et al. (2004). The neoclassical approach limited its view on the nature of the economic agent – homo economicus - to a set of basic assumptions about its behavior. In the second half of the 20th century, the number of phenomena that, in light of contemporary theories, were deemed anomalies could no longer justify ignoring research on the psychology of economic agents. From Simon (1955) to Tversky and Kahneman (1974) and Thaler (1980), behavioral economics has become part of mainstream economic analysis. From the very beginning, experiments played a significant role. Carefully designed experiments helped uncover anomalies as phenomena of perfectly rational behavior when one accounts for so-far-ignored psychological motives. In a meticulous process of ruling out alternatives, experiments are used to identify logical explanations to this day. Behavioral economics has also maintained the neoclassical approach of utility maximization and the search for equilibria. Theoretical work describes behavior positively and extends the standard model with psychological insights. An agent maximizes their utility based on the new assumptions about their information, motivation, and considerations. This thesis combines experimental research with some theory on psychological motivations in moral decision-making and negotiation.

All three chapters in this thesis are situated in the field of behavioral economics. We theorize about the impact of non-standard yet established psychological motives and investigate their effects in traditional economic settings through experiments. Our focus lies on image concerns and social norms in negotiations and moral decision-making. Chapter 2 focuses on nudging opposing social norms and analyzing a possible interaction with social image concerns involving asymmetric information in a (dis)honesty decision. Chapter

3 focuses on social image concerns and their impact on the first-mover preference in negotiations and bargaining. Chapter 4 focuses on self-image concerns and their impact on the first-mover preference in negotiations and bargaining. The overarching topic of all three papers is image concerns (e.g. Bénabou and Tirole (2006), Bodner and Prelec (2003), and Bénabou and Tirole (2011)). The concept of image concerns splits into social and self-image concerns. Social image concerns propose that people are interested in what others think of their motivations. Self-image concerns refer to what an individual thinks of themselves, which assumes that people remember their past actions rather than their true motivation. Image concerns become effective in the decision-making process when either an outside observer or the individual themselves infers their motives from their chosen action. Thereby, even an individual decision becomes a signaling game with either a real or an anticipated outside observer, or with the ever-present inside observer. An action serves as a signal of the individual's inherent preference type. Image concerns contest that people are really interested in the allocation of resources itself. Instead, they propose that people want to convey a beneficial image of their motivation. Examining image concerns gives further insights into the phenomenon of prosocial behavior. Image concerns rely on the concept of social norms. Their close connection explains why social norms enter our investigation of image concerns. An advantageous image must always align with what a relevant peer group identifies as appropriate behavior. Further, we work with a definition of social norms from Bicchieri (2016). A social norm is a behavioral rule that individuals prefer to conform to if they believe that most people in their relevant peer group do so (empirical expectations) and that most people in the same group believe they should conform and may punish deviations (normative expectations). What the individual understands as their relevant peer group depends on their understanding of their identity (Akerlof and Kranton (2000)). The contemporary world offers individuals greater freedom to define who they are, and therefore, identity has become an increasingly important topic. Our origin determines who we are to a lesser extent, while the choices we make in life carry more weight. In such a world where, presumably, belonging depends on signaling an image of who you are and which values you project, signaling might matter more than the actual value itself. All three chapters examine the meaning of image concerns in relation to social norms in traditional economic settings. Chapters 3 and 4 both concern negotiation and bargaining. Chapter 2 looks at asymmetric information and (dis-)honesty decisions. While chapters 2 and 3 focus on social image concerns, chapter 4 deals with self-image concerns.

Chapter 2 (co-authored with Christoph Huber, Annika Nieper, and Timo Promann) focuses on the impact of social norms and social image concerns on (dis)honesty decisions. We experimentally examine their role by manipulating the observability of actions and nudging different norms. If observability of behavior affects the transmission of a social norm and social image concerns are stimulated when behavior is observed, then these mechanisms should amplify honest and dishonest behavior equally. We propose that image concerns can only change behavior in the same direction as the relevant social norm. If there exists a social norm for dishonesty, transparency can lead to adverse behavior through image concerns, just as a social norm for honesty would encourage the opposite. We apply a die-rolling task with a factorial treatment design in an online experiment. To test our proposition, we nudge different norms and stimulate social image concerns. We follow Bicchieri and Dimant (2022) and nudge a social norm by manipulating subjects' empirical expectations. Each participant is grouped with three other participants and receives information about their decision in the die-rolling task. We group those trios depending on their decision and thereby nudge (dis)honesty. On the second dimension, we manipulate social image concerns by varying observability. In one condition, the trio observes the individual's decision in the die-rolling task. In the other, the decision remains hidden.

Our study contributes to the literature on (dis)honest behavior in several ways. First, we add to our understanding of how social norms, along with transparency and observability, can shape (dis)honest behavior. To the best of our knowledge, the present study is the first to explicitly examine a potential interaction effect between these well-established drivers of (dis)honesty and thereby shed light on a potentially damaging effect of increased transparency. We apply peer observation as norm-nudge stimuli (e.g., Bicchieri and Dimant (2022)) and thereby follow recent calls to examine "peer-nudging" as a means of changing social-norm perceptions (Isler and Gächter (2022)) and, eventually, actual behavior. This allows us to identify both empirical and normative expectations as important factors in determining honesty. And indeed, our results show that observing a reference group of peers behaving in a certain way shifts participants' expectations — even in an anonymous online setting. This provides evidence of a norm nudge that can successfully induce behavioral change in ethical decision-making. In a similar vein, with our novel experimental setup in an anonymous online framework, we also contribute to recent discussions on dishonesty in the digital age (Cohn et al. (2022)).

Our data show that social-norm nudges can effectively mitigate dishonest behavior. There is significantly less dishonesty when participants are presented with an honest trio than with a dishonest one. Our rather anonymous online experiment fails to identify any effect of observability on the (dis)honesty decision. To collect empirical and normative expectations, we employ a belief-elicitation task similar to that of Krupka and Weber (2013). Although we observe an increase in lying as we nudge an adverse social norm, normative expectations transmit an unchanged disapproval of lying over all treatments.

Chapter 3 (co-authored with Fanny Schories) investigates the effect of social image concerns on the first-mover preference in negotiations. We ask whether people are less willing to take the potentially advantageous first-mover position when an audience observes their decision. We abstract from complexity by reducing our research to a simplified environment and examine how social image concerns affect the role preferences in a dictator game experiment. In a computerized laboratory experiment, we pair subjects to compete in a second-price auction. Afterwards, a lottery allocates the dictator and recipient roles in an ensuing dictator game. The winner of the auction gains a probability advantage in the lottery to become the dictator. Next, the subjects play a standard one-shot dictator game under one of two treatment conditions. The *public* treatment requires every dictator to stand in front of all participants at the end of a session, while it is publicly announced how much they shared as dictators, and their respective recipients can identify them. We use this design element to stimulate dictators' image concerns. The control condition *private* omits this stage and is equivalent to the usual anonymity protocol in dictator experiments. Finally, we use the auction bids as a measure for the subjects' first-mover preference.

Our research broadly contributes to understanding the phenomenon of prosocial behavior in distributional decisions by investigating the role of social-image concerns. Furthermore, we extend our understanding of the first offer in negotiation by the behavioral economics perspective. In economics, the dictator role is widely accepted as preferred to the recipient role. Our experiment is the first to investigate the intensity of first-mover role-preference in dictator games alongside social-image concerns. We present a potent mechanism to elicit and examine the first-mover preference. The most closely related studies are Dana et al. (2006); Dufwenberg and Muren (2006); Andreoni and Bernheim (2009). Dana et al. (2006) find that subjects are willing to leave money on the table to escape the dictator role. Andreoni and Bernheim (2009) theoretically show how image concerns make dictators more likely to split the surplus equally in a signaling model and report supporting experimental evidence. In contrast, Dufwenberg and Muren (2006) find that audience effects make dictators *less* generous. Regarding these results, one should note that this specific selection of experimental participants was students in an economics lecture. All these studies hint at an effect of the audience on the individual decision. Our study ties the knots and shows how being observed by an audience affects the first-mover preference even in a dictator game.

As hypothesized, we find that dictator offers are higher in the *public* treatment. The main driving factors are one's own fairness concerns and awareness of one's image concerns. Correspondingly, the willingness to pay to become the dictator is significantly lower in the *public* treatment compared to the *private* treatment. Here again, one's own fairness and awareness of one's own image concern explain the results. Additionally, in the *private* treatment, high expectations about others' fairness decrease auction bids. High expectations increase the value of the recipient role, thereby decreasing bids. With the public announcement, this effect vanishes as overall expected generosity rises.

Chapter 4 investigates the effect of self-image concerns on the first-mover preference in negotiations and thereby also offers a behavioral economics perspective on the first-offer dilemma. We use the novel preference-elicitation mechanism first introduced by Litsios and Schories (2024) to understand the impact of self-image concerns on the first-mover preference. We conduct an online experiment in which we auction the dictator role of an ensuing game in a second-price auction. We understand changes in bidding behavior as shifts in subjects' first-mover preferences. We control self-image concerns by offering participants two potential self-image concern avoidance strategies. This allows us to measure their impact on the first-mover preference between treatments. The two strategies we use are belief manipulation (Bénabou and Tirole (2016) and Gino et al. (2016)) and strategic ignorance (Dana et al. (2007)). Belief manipulation enables individuals to behave selfishly while maintaining a positive self-image by strategically selecting beliefs about appropriate behavior. Our treatments vary in the ability to strategically choose beliefs by varying the timing of information provision. In one treatment, beliefs are elicited before subjects know whether they will participate further in the main task. In the other treatment, beliefs are elicited after subjects learned about their participation in the main task. This procedure is inspired by Bicchieri et al. (2023). Strategic ignorance allows individuals to behave selfishly while maintaining a positive self-image by deliberately ignoring the consequences of their actions. Our treatments vary in their ability to maintain strategic ignorance by introducing uncertainty about the recipient's payoff consequences and placing the decision to costlessly reveal the true state in the subjects' hands. This procedure is inspired by Dana et al. (2007).

Our study contributes to the economic literature in several ways. First, we extend research on prosocial behavior by shedding light on the role of self-image concerns in distributional decisions. Our findings support Dana et al. (2007) and others, who argue that prosocial behavior is context-dependent and driven by a more complex psychological motive than a simple preference for fairness. Second, we apply this psychological motive to the archetype of economic interaction – bargaining and negotiation. Especially, we examine another behavioral layer of the first-offer dilemma. Third, by applying the avoidance strategies of belief manipulation and strategic ignorance, we further investigate how self-image concerns can be manipulated, while fourth, we deepen our understanding of the connection between self-image concerns and social norms. Finally, our research adds to our understanding of the benefits and shortcomings of online experiments.

Our results give rather weak or no support for our hypothesis about the importance of self-image concerns for the first-mover preference. On the level of the dictator game, our two manipulation strategies show an impact. First, we find mild evidence that belief manipulation increases the first-mover preference. Subjects' beliefs shift in a pessimistic direction, and we observe more selfish behavior. Average auction bids do not differ significantly across treatments, and despite an increased number of selfish subjects, bid distributions remain unchanged. Average bids and distributions of selfish subjects do not differ between treatments. This suggests that belief manipulators have preferences similar to those of purely selfish subjects; however, given the small sample sizes, the increase in the number of selfish subjects does not translate into a significant shift in auction bids in the global comparison of treatments. Second, our attempt to isolate the effect of strategic ignorance was unsuccessful because a design flaw introduced additional variation between the treatments, making it difficult to attribute changes in behavior solely to strategic ignorance. Anyhow, the combination of effects in the strategic ignorance treatments almost entirely deteriorates pro-social behavior and highlights the importance of self-image concerns.

Overall, the three papers deepen our understanding about the role of image concerns in prosocial behavior and its connection to social norms. Further, we propose a novel mechanism to evaluate the first-mover preference, test strategies to manipulate self-image concerns and gain new insights into online experimentation in the behavioral economic field. Chapter 2 focuses on nudging opposing social norms and analyzing a possible interaction with social image concerns involving asymmetric information in a truth-telling decision. Chapter 3 focuses on social image concerns and their impact on the first-mover preference in negotiations and bargaining. Chapter 4 focuses on self-image concerns and their impact on the first-mover preference in negotiations and bargaining.

Bibliography

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, 115(3):715–753.

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.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678.

Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The quarterly journal of economics*, 126(2):805–855.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms.* Oxford University Press.

Bicchieri, C. and Dimant, E. (2022). Nudging with care: The risks and benefits of social information. *Public choice*, 191(3):443–464.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2023). 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.

Bodner, R. and Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. *The psychology of economic decisions*, 1(105):26.

Camerer, C. F., Loewenstein, G., and Rabin, M. (2004). *Advances in behavioral economics*. Princeton university press.

Cohn, A., Gesche, T., and Maréchal, M. A. (2022). Honesty in the digital age. *Management Science*, 68(2):827–845.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and human decision Processes*, 100(2):193–201.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Dufwenberg, M. and Muren, A. (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization*, 61(1):42–49.

Gino, F., Norton, M. I., and Weber, R. A. (2016). Motivated bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives*, 30(3):189–212.

Isler, O. and Gächter, S. (2022). Conforming with peers in honesty and cooperation. *Journal of Economic Behavior & Organization*, 195:75–86.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Litsios, C. and Schories, F. (2024). Making the first offer in negotiation – explaining role preferences by social image concerns in the dictator game.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, pages 99–118.

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, 1(1):39–60.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185(4157):1124–1131.

Chapter 2

On Social Norms and Observability in (Dis)honest Behavior

Authors: Christos Litsios, Christoph Huber, Annika Nieper, Timo Promann

Abstract

Transparency and observability have been shown to foster ethical decision-making as people tend to comply with an underlying norm for honesty. However, in situations implying a social norm for *dis*honesty, this might be different. In a die-rolling experiment, we investigate whether observability can also have detrimental effects. We thus introduce a norm nudge toward honesty or dishonesty and make participants' decisions observable and open to the judgement of other participants in order to manipulate the observability of people's decisions as well as the underlying social norm. We find that a nudge toward honesty indeed increases the level of honesty, suggesting that such a norm nudge can successfully induce behavioral change. Our introduction of social image concerns via observability, however, does not affect honesty and does not interact with our norm nudge.

Keywords: lying, cheating, social norms, image concerns, nudging, behavioral change

JEL Classification: C91, C92, D01, D91, M14

2.1 Introduction

People frequently engage in dishonest behavior, which can—depending on the context—entail large costs to individuals, organizations, and society more broadly. For policy design that effectively fosters honesty, it is thus pivotal to explore and identify relevant behavioral mechanisms. In the present study, in particular, we focus on social norms, people's preferences for being seen as honest, and the way these concepts interact. For that purpose, we set up an experiment in which we vary whether information about others' reporting represents team members as being honest or dishonest, and whether own reporting behavior is observable to team members.

A recent meta-analysis of experimental studies on (dis)honesty concluded that people tend to have a preference for being honest and for *being seen as honest*, i.e., people have image concerns (Abeler et al., 2019). Previous research shows that *norms* also play an important role in people's prosocial behavior (e.g., Bicchieri, 2016; Bicchieri et al., 2022; Cialdini et al., 1990; Krupka and Weber, 2013) and that "whenever individuals believe they are expected by their group ... to behave according to a given standard, and also expect the norm to be generally followed, they usually comply." (Bicchieri and Muldoon, 2011).

In many decision situations, a norm for honesty prevails. If this is the case, then the previous literature suggests that observability—and thus image concerns—encourages honesty. However, it is easy to conceive of decision situations in which there appears to be a norm for *dis*honesty—that is, in which one is expected to comply with others' *dis*honest behavior. In those cases, we propose that it is reasonable to expect observability to encourage *dis*honesty instead. Consider the example of fraudulent practices in organizations, for instance, where there prevails a social norm—as well as an expectation—for unethical behavior internally, which might differ from external norms. Accordingly, this mechanism may result in corruption, fraud, and related malfeasance (e.g., Slemrod, 2007; Dyck et al., 2023). Similar applications include people observing the behavior of any reference group, their peer group, or even their leaders, and thus learn about the prevailing norm, which might just as well not be ethical or honest in nature (e.g., Acemoglu and Jackson, 2015; Ajzenman, 2021).[1] Examining the role of observability under different social norms is thus key to advancing our understanding of (dis)honest behavior more generally, and it directly relates to topics that span important domains in economics and the social sciences.

Previous research has shown that targeting people's concern for being seen as honest can increase honesty (Bo et al., 2015; Bodenschatz and Irlenbusch, 2019). People cheat less in an experimental die-rolling task if they are observed by others (e.g., Gneezy et al., 2018; Fries et al., 2021; Bašić and Quercia, 2022). Nevertheless, increasing people's concern for how others view them can also have no effect on honesty; it can even increase dishonesty. This can be seen, for instance, when one person, who benefits from mutual dishonesty, can observe another's decision (Weisel and Shalvi, 2015; also see Kocher et al., 2018). These mixed findings might result from differing expectations of others to behave (dis)honestly based on the situation, as different situations can evoke different social norms (Akerlof and Kranton, 2005). Observing others' actual behavior, in particular, may alter the empirical expectations on what constitutes the social norm in a given context (Bicchieri, 2016). A number of previous studies have used such information provision as a social norm nudge—that is, to evoke perceptions of "good" and "bad" norms (Bicchieri and Dimant, 2022; see Bicchieri and Xiao, 2009, for example, in Dictator games; and Isler and Gächter, 2022, in honesty and cooperation).

[1] If there is widespread dishonesty in a specific context and if that is widely known or perceived, then one might infer that there is a social norm for dishonesty, which may affect others' behavior. Important applications include tax evasion (Slemrod, 2007), insider trading (Acharya and Johnson, 2010), academic cheating (Jensen et al., 2002), or misrepresenting information in politics (Swire-Thompson et al., 2020), among other examples.

The present study stands out in this body of research because it applies at least two novel design elements: first, we introduce the interaction between information provision (observing other participants' decisions) and observability (one's decision is observed by other participants); second, both information provision and observability relate to the same group of peers—i.e., the participants one observes in a first stage are precisely those who then observe one's own decision in a second stage. With this novel setup and the introduction of an additional stage in which the same peers provide feedback on one's decision, we also introduce an element of accountability that is not present in previous work.

In this study, we experimentally manipulate observability and examine the role of social image concerns in (dis)honesty decisions under different social norms. If norms shape behavior and observability affects adherence to norms, then the introduction of observability should increase honest and dishonest behavior alike, in line with the respective norms. We thus propose that the ability of image concerns to increase honesty relies on a social norm for honesty. If there prevails a social norm for dishonesty instead, image concerns might even have a backfiring effect. We test this proposition in an experimental die-rolling task with a factorial treatment design, in which we manipulate people's perception of the prevalent social norm, as well as their social image concerns. Following Bicchieri and Dimant (2022), we induce a social norm nudge by manipulating subjects' empirical expectations on how other people would act in a comparable situation. For that, each participant is grouped with three other participants and then presented with (social) information about those participants' decisions. To vary subjects' social image concerns, we manipulate the observability and accountability of their decisions: in one set of conditions, a participant's decision remains unobserved by others, while in another set of conditions their decision is shown precisely to the reference group of participants used to induce the norm nudge. Note, however, that in contrast to previous literature, neither the framing (e.g., Dimant et al., 2020) nor the rules of the game (e.g., Serdarevic, 2021) differ between treatments.

Our results demonstrate that social norm nudges can be effective in mitigating dishonest behavior. We find significantly less dishonesty when participants are presented with others' honest behavior, compared to when they are presented with others' dishonest behavior. In our anonymous online setting, we find no effect of observability on people's (dis)honesty decisions. Importantly, our design also allows us to elicit empirical and normative expectations on the inherent social norm in each individual treatment condition. For that purpose, we implement a belief elicitation inspired by the Krupka and Weber (2013) paradigm: Participants are asked to assess the social appropriateness of lying in the die-roll task, as well as the prevalence of actual lying, and are incentivized in a coordination game setup; specifically, they receive a bonus payment if they anticipate the most frequently given assessment (for a more general discussion on the use of coordination games as an elicitation method, see Schmidt et al., 2022). Although we do observe a significant uptick in lying when norms shift toward dishonesty, we note that lying is, on average, not seen as socially appropriate in any of our treatments.

This study contributes to the literature on dishonest behavior in a number of important ways. First, we add to our understanding of how social norms, as well as transparency and observability, can shape (dis)honest behavior. To the best of our knowledge, the present study is the first to explicitly look into a potential interaction effect between these well-established drivers of (dis)honesty and thereby shed light on a potential damaging effect of increasing observability. We apply peer observation as norm-nudge stimuli (e.g., Bicchieri and Dimant, 2022) and thereby follow recent calls for examining "peer-nudging" as a means of changing social-norm perceptions (Isler and Gächter, 2022), and, eventually, actual behavior.[2] This allows us to identify both empirical and normative expectations as important factors in determining honesty. And, indeed,

[2]A related yet distinct concept is "meta-nudging" (Dimant and Gesche, 2021; Dimant and Shalvi, 2022), which aims to achieve behavioral change by nudging people indirectly via "social influencers"; that is, those with the ability to enforce other's behavior and norm adherence. "Peer-nudging," by contrast,

our results show that observing a reference group of peers behaving in a certain way does shift participants' expectations—even in an anonymous online setting. This provides evidence of a norm nudge that can successfully induce behavioral change in ethical decision-making. In a similar vein, with our novel experimental setup in an anonymous online framework, we also contribute to recent discussions on dishonesty in the digital age (Cohn et al., 2022).

The remainder of this paper is organized as follows. In Section 2.2, we review relevant literature related to our research questions and outline the hypotheses to be tested. Section 2.3 describes our experimental design and procedure. In Section 2.4, we present our main results. Section 2.5 concludes.

2.2 Related Literature and Hypotheses

Ever since Becker (1968) published his rational crime model, economists and other social scientists have studied in what situations and with what motives people behave honestly. Of greatest interest are situations in which honest behavior departs from the standard economic model's prediction—that is, when lying would maximize payoffs. In their seminal work, Fischbacher and Föllmi-Heusi (2013, FFH) reported 39% of experimental participants to be "fully honest." Much of the following economics literature on lying and dishonesty uses, adapts, and extends their clever but simple incentivized experiment: Participants privately roll a six-sided die, are asked to memorize the number that came up, and are subsequently asked to report this number, where different numbers are associated with different monetary payoffs. Lying in this task is defined as a participant reporting a number different than the one they rolled and is measured by the overall distribution of reports in lieu of observing individuals' die rolls to maintain anonymity. This and a number of related experiments have identified lying costs, i.e., intrinsic costs derived from deviating from truth-telling, which foster lying aversion (e.g., Ellingsen and Johannesson, 2004; Kartik, 2009; Gneezy et al., 2013, 2018; Abeler et al., 2019). In a meta study of 90 different experiments using variants of the FFH design, Abeler et al. (2019) concluded that overall, people "lie surprisingly little" and named "a preference for being seen as honest and a preference for being honest" as the primary motivations (Abeler et al., 2019, p. 1115).

Several researchers have taken up the notion that people wish to be perceived as honest and have extended a simple model of lying costs to incorporate what Abeler et al. (2019) termed a "reputation for honesty"—i.e., social image (or social identity) concerns (e.g., Dufwenberg and Dufwenberg, 2018; Khalmetski and Sliwka, 2019). More generally, one can distinguish self-image concerns, when behavior remains unobserved by others; social image concerns, when behavior is observed by others but payouts remain independent; and reputational concerns, when behavior is observed by others and entails interrelated payouts (see Bolton et al., 2021). The main conceptional difference between image concerns and reputational concerns is thus rooted in the payoff independence between decision maker and observer(s). Gneezy et al. (2018) compared lying behavior in an FFH-type experiment between an observed and a non-observed condition. In the observed condition, decisions were made on a computer and could be observed by the experimenter; in the non-observed condition, however, sealed envelopes were used to ensure anonymity. The researchers found more prevalent lying in the non-observed condition than in the observed one, suggesting that social image concerns are an important determinant in lying behavior. In another die-rolling task, Fries et al. (2021) implemented different levels of observability of participants' die rolls and reports. They found that an increase in the die roll's observability, in particular, could facilitate honesty, catering to social signaling motivations, as

refers to nudging people through social information about their peers.

perfectly identifying liars becomes possible. While most of the previous work in this direction has focused on people giving up monetary rewards in order to appear honest, Barron et al. (2021) observed a willingness to lie for a desirable (social) image. Bolton et al.'s (2021) results from a series of dictator game experiments also provide valuable insights for the research questions under investigation in the present study: They reported that observability—and thereby social image concerns—could only have little or even negative effects in certain situations.

Social norms are closely related to social image concerns and present another prominent for (dis)honest behavior. The underlying idea in this line of research is that utility also depends on the perceived distribution of lying costs, in the society or among a particular reference group (e.g., Weibull and Villa, 2005; Gibson et al., 2013). Social norms, more generally, are rules that prescribe appropriate behavior and build upon an individual's expectations about how others behave (empirical expectations) and about how others believe one ought to behave (normative expectations; Bicchieri, 2016).[3] In this regard, a few studies have experimentally induced norm-nudge stimuli, that is, interventions that aim to change aforementioned expectations (Bicchieri and Dimant, 2022). In a recent study, Dimant et al. (2020), for example, sought to achieve behavioral change by norm-nudging (dis)honesty. They made use of framing effects in presenting participants with information about a majority of participants having been honest or a minority of participants having been dishonest (or the analogue normative information, depending on the treatment) before asking participants to make a decision themselves. Nevertheless, they reported null effects, as the intervention did not achieve a shift in participants' perception of the prevailing social norm. In a closely related study, Serdarevic (2021) introduced a different nudge: Societal expectations were varied by adding information on the expected truthfulness in an FFH-type decision task (participants were informed that they either "have to report truthfully," "do not have to report truthfully," or no additional information was given in a control condition). Serdarevic reported significantly more dishonesty among participants encouraged to misreport, and indeed, she identified a shift in what is seen as socially appropriate to be driving this behavioral change.[4] Likewise, empirical data from Ajzenman (2021) highlights the importance of social norms in (dis)honest behavior: He found an association between an uptake in students' cheating and revelations of corruption among local officials. Several recent studies have also investigated the role of social proximity with respect to social norm compliance: Dimant (2019), for example, found that social proximity amplifies the contagion of anti-social behavior, in particular; and Bicchieri et al. (2022) also identified social proximity as a key ingredient for norm compliance among peers.

While the literature heretofore discussed focuses on how people's individual willingness to lie is shaped by intrinsic lying costs, social image concerns, and social norms, several studies have examined lying behavior in a social context more explicitly by investigating group behavior. In the context of corruption, Weisel and Shalvi (2015) introduced a dyadic game in which two players sequentially play a die-rolling task. By varying the extent to which the two players' payoffs were aligned, they found vast dishonesty with perfectly aligned incentives at almost 50%

[3]Some of the literature we are citing uses the expressions empirical and normative expectations when referring to what others call descriptive (how others behave) and injunctive (how one believes one ought to behave) norms. For consistency, in this paper we only refer to empirical expectations and normative expectations when referring to these concepts (also see Bicchieri and Dimant, 2022, for example).

[4]In an experiment on corruption behavior, Köbis et al. (2015) manipulated empirical expectations in a positive or negative way by priming participants with information suggesting that either "almost nobody" or "almost everybody" made a corrupt decision. Similarly, participants in Lois and Wessa (2020) received false feedback about the average level of (dis)honest behavior to induce empirical and normative expectations of the social norm in each of three different treatments. While both of these studies found less ethical behavior in a pro-corruption norm condition and a condition suggesting high cheating among peers, respectively, we regard these methods as (at least borderline) deceptive, as the social information presented to participants is fictitious.

higher levels than in individual decisions. Moreover, Kocher et al. (2018) observed significantly more lying when decisions are taken as a group of three than as individuals, even without payoff commonality among group members. Importantly, they also identified a an upward shift in group members' expectations about others' lying behavior through communication, which may have additionally facilitated their own dishonesty. These studies highlight that, depending on the particular situational context, people might be even more inclined to lie in the presence of others.

Taken together, previous literature has shown that both social image concerns and social norms can have a positive effect on honesty, whereas group settings, in which people's payoffs depend on one another, might increase cheating. Nevertheless, cheating can be prevalent in group decisions, even when payoffs are not interrelated, such as when common expectations about others deteriorate toward dishonesty. We hypothesize that social image concerns do increase honesty when the norm is to be honest, but that such concerns can have a backfiring effect when there is a norm for dishonesty—social image concerns might lead to stronger conformity to the underlying social norm and thereby encourage honesty *or* dishonesty depending on the situational context. While, in principle, these propositions do not rely on social proximity, we expect them to be stronger with a shorter social distance to the reference group. Hence, we postulate the following three main hypotheses to be tested:

- H1 People behave more honestly when there is a social norm for honesty than when there is a social norm for dishonesty.
- H2 People's decisions to behave honestly or dishonestly are affected by social image concerns.
- H3 Social image concerns and social norms have an interactive effect on honesty. Specifically, social image concerns increase honesty under an honest norm and decrease honesty under a dishonest norm.

2.3 Experimental Design and Implementation

We examine (dis)honest behavior and test our hypotheses using a die-rolling game (Fischbacher and Föllmi-Heusi, 2013), in which behavior has been shown to correlate with various instances of unethical behavior in naturally occurring settings (e.g., Dai et al., 2018, Hanna and Wang, 2017). In this game, participants roll a computerized fair six-sided die, which can be any integer between 1 and 6: $d \in \{$"⚀","⚁","⚂","⚃","⚄","⚅"$\}$. The computerized die roll is implemented by presenting a randomly chosen video recording of an actual die roll on participants' screens. Each participant is then asked to enter their rolled number, making each individual decision fully observable by the experimenter in all treatments (Kocher et al., 2018). Only one of the six possible outcomes yields a bonus payment for participants, such that they have to make a binary decision between lying or not lying. Thus, if participants report having rolled "⚃", they earn a bonus of £1.50 on top of the equally sized reward for participation. If they report having rolled any other number (1, 2, 3, 5, or 6), they earn no bonus. This design choice rules out partial lying (Kajackaite and Gneezy, 2017), which might be affected by observability (Gneezy et al., 2018; Abeler et al., 2019).[5] In comparison to a binary coin toss, this design also increases the

[5] Gneezy et al. (2018) and Abeler et al. (2019) found that when participants can lie partially (i.e., report a payoff-enhancing but not the payoff-maximizing outcome), they do so less in observable compared to non-observable conditions; full lying (i.e., reporting the payoff-maximizing outcome), in contrast, is hardly affected by choices being observable to the experimenter. Crede and von Bieberstein (2020) nevertheless found considerably less lying when participants were explicitly made aware of their choices being observable by the experimenter. As our experimental instructions do not explicitly mention that the experimenter is able to track outcomes (in line with Gneezy et al. (2018) and Abeler et al. (2019)), and as partial lying is not possible in our experiment setup, we expect potential effects of experimenter observability to be

proportion of participants with an incentive to lie, as the ex ante probability of missing out on the payoff-maximizing outcome (“⚅”) is $\frac{5}{6} = 83.3\%$. At the same time, participants are generally familiar with six-sided dice such that the randomization device and corresponding probabilities are intuitive and easy to understand.

2.3.1 Experimental Treatments and Procedure

We employ a 2 × 2 between-subjects design, in which we vary the observability of a subject's decision and introduce norm nudges to foster an honest or dishonest social norm; see Table 2.1. Each participant is randomly assigned to one of the four treatments.

Table 2.1: Treatment overview

		Observability	
		Public	Private
Norm nudge	Honest	*honest-public* (T1)	*honest-private* (T2)
	Dishonest	*dishonest-public* (T3)	*dishonest-private* (T4)

To operationalize our treatment variations, we apply the following experimental procedure. Before an experimental session begins, each participant is randomly assigned either role A or role B. The experiment then consists of two stages. In the first stage, only A players play the die-rolling game, as described above. Each A player can then be categorized as honest or dishonest: If an A player reports having rolled “⚅” while their die roll yielded a different number, they are categorized as *dishonest*; if they correctly report the number of pips on their die roll, they are categorized as *honest*. If an A player's report is different from “⚅” but also different from the outcome of their die roll, they are excluded from further stages of the study, as their behavior could be seen as dishonest but they did not lie to increase their monetary payoff. Once a sufficient number of A players have made their decisions, we form groups of three consisting of either three honest or three dishonest A players.

In the second stage, B players are asked to complete the same die-rolling game. They are informed that they have been matched with three A players to form a team, but the exact matching mechanism is not revealed.[6] B players in all treatments then see their team's die rolls *and* their reports, such that there is no uncertainty about the extent of (dis)honesty in the reference group.[7] This procedure serves as our social norm manipulation, nudging participants' perceptions of the social norm toward honesty or dishonesty. As the experiment is conducted online, each

negligible and constant across treatments.

[6]Presenting a non-representative sample without mentioning the precise sample selection follows the goal of inducing the perception of a social norm for honesty or dishonesty. On a general level, the omission of information is not necessarily regarded as deception in experimental economics (e.g. Hey, 1998; Hertwig and Ortmann, 2008; Wilson, 2016) and a recent survey among student participants and experimental economics researchers showed that presenting a non-representative sample was only regarded as rather deceptive by around 20% of researchers and by even fewer students (mean rating 3.76 on a seven-point Likert scale); the majority of researchers regarded this method as appropriate (mean rating 4.76 on a seven-point Likert scale; see Charness et al., 2022). In fact, our use of a subtle and neutral phrasing (e.g, “You are now matched with three A players.”) is in line with the proposed alternatives to presenting a non-representative sample as reported in Charness et al. (2022).

[7]Every group of three A players is separately matched with six B players to increase monetary efficiency because of budgetary constraints.

participant naturally has a large spacial and social distance to other participants. To reduce the perceived social distance, they are informed about forming a "team" with other participants, which are each represented by a self-selected avatar (from a set of possible gender-neutral avatar choices; see Abraham et al., 2023, for example).

In the set of *public* treatments, we then seek to induce observability. The outcomes of B players' die rolls and their reported numbers are thus explicitly communicated to their team. In light of the high level of subject anonymity in an online experiment, we introduce an additional feature to strengthen this condition: After B players have made their decisions, the same group of A players is asked to provide each matched B player with feedback about their behavior in the game (without any monetary consequences).[8] Players are informed about this procedure in the experimental instructions. In the set of *private* treatments, in contrast, A players are not informed about B players' behavior and do not provide any feedback, instead moving directly to a concluding survey. Figure 2.1 provides an overview of the experimental procedure.

[8]In particular, the feedback consists of two parts: a numerical rating on a scale from "very negative" (-3) to "very positive" (3) and a verbal response in the form of an adjective describing how participants perceived the B player's behavior. In addition, each A player has the opportunity to revise both their numerical rating and their verbal response after being informed about the feedback from the other two A players in their group. We thereby seek to mimic a simplistic group discussion involving potential gossip concerning B players' behavior to further increase social image concerns (see, for example, Bénabou et al., 2020).

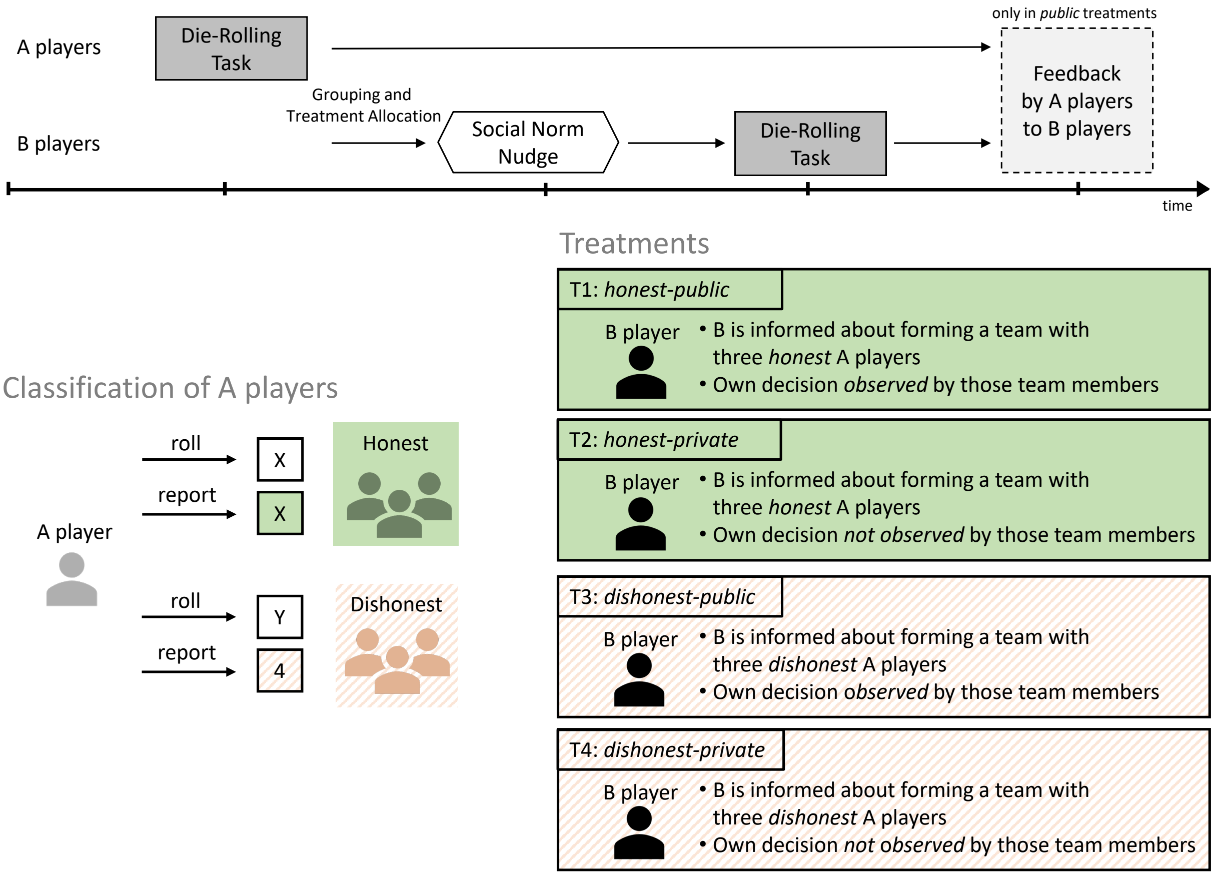


Figure 2.1: **Overview of the experimental procedure and design.** This figure schematically outlines the experimental procedure, classification of A players into honest and dishonest players, and the treatment allocation for B players. In the die-rolling task, participants first see a video of a six-sided die roll and are then asked to report the rolled number. Regardless of the actual die roll, they receive a bonus payment if they report the number 4. A players are then classified depending on their report: if they report the number of their die roll, they are classified as honest (with $X \in \{1, 2, 3, 4, 5, 6\}$); if they report 4 instead of the number of their die roll, they are classified as dishonest (with $Y \in X \setminus \{4\} = \{1, 2, 3, 5, 6\}$). In the social-norm-nudge stage, B players are then presented with the die rolls and reports of either three honest A players (T1 and T2) or three dishonest A players (T3 and T4) before going through the die-rolling task themselves. Note that the 'Feedback by A to B players' stage only appears in the two *public* treatments (T1 and T3).

Following B players' decisions in the die-rolling game but before receiving A players' feedback, we elicit their personal norms (Bašić and Verrina, 2021; Bicchieri and Chavez, 2010; Bicchieri and Xiao, 2009) and their normative and empirical expectations (Krupka and Weber, 2013; in that order). Participants are asked how socially appropriate they believe reporting "⚃" to be when a different number was rolled (personal norm) and how socially appropriate they believe most others find this behavior (normative expectation). Both variables are measured on a six-point scale from "very socially inappropriate" to "very socially appropriate." We then elicit empirical expectations by asking, "What percentage of participants do you think reports '⚃' when a different number ... was rolled ... ?" Correct responses concerning normative and empirical expectations, respectively, are rewarded with £0.25. Finally, we measure participants' social and honest image concerns with seven additional survey items.[9]

2.3.2 Experimental Implementation

The experiment was programmed and conducted using oTree (Chen et al., 2016). Participants were recruited via *Prolific* (see, for example, Palan and Schitter, 2018). Participants were paid a participation fee of £1.50 and were able to earn a bonus payment of up to £2.00 depending on their decisions in the experiment, as outlined in the previous section. Participants needed to be at least 18 years old, speak English fluently, have an approval rate on Prolific of at least 90%, and have previously participated in at least five studies.

We recruited a total of 1,629 participants. As we are mainly interested in B players' behavior but required a sufficiently number of A players acting as a reference group and providing feedback to B players in the public treatments, we randomly assigned role A with a 1/3 probability, and role B with a 2/3 probability. This procedure resulted in a sample of 543 A players and 1,086 B players. The 543 A players were then allocated into 181 groups of three players each, either all *honest* or all *dishonest*. Each of these groups was matched with six B players; that is, each group acted as the reference group for six different B players.[10] Participants who did not report any number in the die-rolling game were excluded. Additionally, we applied the following preregistered exclusion criteria: participants who misreported the number that they saw but did not report "⚃", participants who were timed out prior to reporting the die roll,[11] and participants who saw the computerized die roll of "⚃" were excluded from the analysis (since those participants had no incentive to lie).[12] This led to a final sample of 1,192 participants, 409 in the role of A players and 783 in the role of B players.

2.4 Results

Overall, 38% of participants were female, the average age was 29 years, they earned a median income of about US$15,000 per year, about half of all participants were students, and the majority was European (22.6% British, 18.2% Portuguese, 11.2% Polish, and 7.5% Italian). The experiment had a median duration of 22 minutes, and participants earned an average of £2.24 (Std. dev: £0.75). Table A1 in the Appendix summarizes the distribution of participant

[9]The survey items on participants' social image concerns are "I was concerned about what others think about me," "it was important to me that my team members would perceive me in a positive way," "it was important to me that my team members would accept me," and "I thought about what information my team members might share about me to another person." The survey items on their honest image are "I wanted others to think I am a person who tells the truth," "I wanted others to think I am a person who does not misrepresent facts," and "I wanted others to think I am a person who does not lie." All items are measured on a seven-point Likert scale from "totally disagree" to "totally agree" (see Wu et al., 2015).

[10]Note that B players, on the other hand, were informed that they were matched with three A players to form a team. With this setup, each group of three A players was separately shown to six B players and acted as their reference group, and each group of three A players in the *public* treatment gave feedback to six B players.

[11]Because the experiment is interactive, all participants in our experiments are asked to provide timely responses such that their respective partners in the experiment do not have excessive wait times. We used pilot data to establish an average participation time. Participants automatically proceeded to the next page when they took too long to respond but thereby forewent any bonus payment.

[12]In addition, we excluded 99 observations from participants who took part in two sessions due to a technical error. In those cases, only data from one's second participation was excluded.

demographics across treatments.

In this section, we will first investigate whether our novel treatment variations yield their intended effect. We will thus examine to what extent the induced social norm nudge affects normative and empirical expectations, as well as personal norms (section 2.4.1), and further look into differences in social image concerns between the two observability treatments, public and private (section 2.4.2). After having established the mechanics of our treatment variations, we will go on to analyze the extent to which they affect lying behavior (section 2.4.3). We differentiate between *honest* and *dishonest* on the social-norm-nudge dimension and *private* and *public* on the observability dimension. As outlined in previous sections, combining the two dimensions results in four treatments: *honest-public* (T1), *honest-private* (T2), *dishonest-public* (T3), and *dishonest-private* (T4).

2.4.1 Social Norm Nudge

For the social-norm-nudge dimension, B players are matched with three honest or three dishonest A players in the *honest* and *dishonest* condition, respectively. After observing those three A players' decisions in the die-rolling game, B players conducted this task themselves. The aim of these treatment variations is to change participants' perceptions of social norms—that is, their understanding of what one ought to do and their expectations of what others do. We thus elicited participants' normative and empirical expectations of the social norm using an incentivized coordination task (Krupka and Weber, 2013). Assuming our treatment variations are successful in manipulating social norms, we expect a shift toward dishonesty in the *dishonesty* nudge treatment in comparison to the *honesty* nudge treatment because participants would perceive lying as more socially appropriate and expect others to lie. We also elicited participants' personal norms (Bašić and Verrina, 2021) to enable us to distinguish a shift in societal expectations from a shift in personal convictions.

To elicit normative expectations, we ask participants, "How socially appropriate do you think most others find reporting '⚃' when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in?" on a six-point scale from "very socially inappropriate" to "'very socially appropriate," coded from -1 to 1 in equal distances. To elicit empirical expectations, we ask, "What percentage of participants do you think reports '⚃' when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in?" The mean answers within all treatments are shown in Figure 2.2 for normative and empirical expectations. The two graphs paint a similar and convincing picture. First, we observe a significant shift toward dishonesty in the set of *dishonesty* treatments, confirming the effectiveness of our treatment manipulation: in the dishonesty treatments (T3 & T4), lying is perceived to be significantly more socially appropriate (public: $p < 0.001$, private: $p < 0.001$; Mann-Whitney U test) and participants expect a lower percentage of people to report truthfully (public: $p < 0.001$, private: $p < 0.001$; t-test). This is also reflected in an analysis of the feedback from A players to B players about their behavior: Lying in the honest treatments evokes strongly negative feedback, while lying in the dishonest treatments has no (numerical feedback) or less negative (verbal feedback) repercussions (see Figure A1 in the Appendix). Another important observation we can see from these results on social norms, however, is that lying is never perceived as socially appropriate. Nevertheless, after a dishonesty nudge, lying on average is perceived as neither socially appropriate nor socially inappropriate, but rather neutral. Finally, it is reassuring that our norm-nudge treatments are indeed effective in shifting normative and empirical expectations, while they remain unaffected by varying observability, as we find no differences between the respective *public* and *private* treatment conditions. Linear regression estimates confirm these initial results (see Table 2.2).[13]

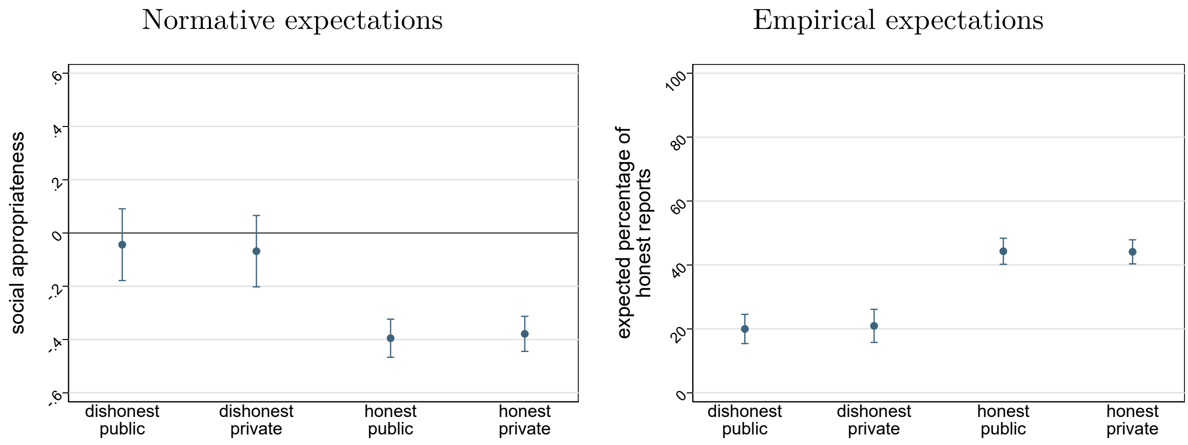


Figure 2.2: **Normative and empirical expectations across treatments.** This figure presents the mean values and 95% confidence intervals of participants' normative expectations (left panel) and their empirical expectations (right panel) across all four treatments. Normative expectations refers to the social appropriateness of a given action from "very socially inappropriate (-1) to "very socially appropriate" (1). Empirical expectations refers to the expected percentage of honest reports in a given situation. Note that we elicited the expected percentage of dishonest reports but present the expected percentage of honest reports ($100 - E(\textit{percentage of dishonest reports})$) for comparability.

To elicit participants' personal norms, we asked, "How socially appropriate do you find reporting '⚃' when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in?" on a six-point scale from "very socially inappropriate" to "'very socially appropriate," which was coded from -1 to 1 in equal distances. While the mean values of public and private conditions paired with the dishonesty nudge are almost identical, the respective means in honesty-nudge treatments differ slightly and are both lower than in the dishonesty treatments. Nevertheless, none of these differences are found to be statistically significant after administering non-parametric Mann-Whitney U tests (all $p > 0.100$), whereas linear regression estimates reveal weak evidence for a dishonesty shift with regard to participants' personal norms. As shown in Column (5), the difference between honest and dishonest nudge is significant at the 10% level ($p = 0.059$), pooling private and public, but when disentangling the two, as seen in Column (6), any significance disappears. Comparing these results with the ones for *social* norms, we find almost identical values with regard to appropriateness of lying in the honesty conditions, whereas there is a considerably larger shift toward approval of dishonesty in social norms than in personal norms. This result confirms prior evidence by Bicchieri et al. (2022), in that empirical and normative expectations can be quite malleable with respect to being exposed to peer behavior and observability, whereas personal norms are usually not as easily swayed.

2.4.2 Observability

In our second treatment variation, we manipulated whether B players' decisions are observable by the respective A players to induce social image concerns. To test the effectiveness of this treatment variation, we elicited two sets of survey questions. The first set measures the degree to which

[13]Observations for normative expectations, empirical expectations, and personal norms differ slightly, as a small number of participants dropped out before completing all survey pages.

Table 2.2: **Linear regression: Normative expectations, empirical expectations, and personal norms.** This table shows the estimated coefficients from linear regressions of normative expectations (1, 2), empirical expectations (3, 4), and personal norms (5, 6) on binary variables indicating the particular treatment. Normative expectations refers to a given action's social appropriateness. Empirical expectations refers to the expected percentage of honest reports in a given situation. Personal norms refers to participants' personal views on a given action's social appropriateness. "Private" takes the value 1 for the set of *private* treatments (T2 and T4) and 0 otherwise; "Honest" takes the value 1 for the set of *honest* treatments (T1 and T2) and 0 otherwise. No controls were included. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Normative		Empirical		Personal	
	(1)	(2)	(3)	(4)	(5)	(6)
Private	0.006	−0.024	−0.121	−0.960	0.042	0.004
	(0.044)	(0.087)	(2.271)	(4.445)	(0.049)	(0.093)
Honest	−0.332***	−0.351***	−23.777***	−24.312***	−0.103*	−0.127
	(0.050)	(0.069)	(2.580)	(3.547)	(0.055)	(0.078)
Private × Honest		0.041		1.136		0.051
		(0.101)		(5.172)		(0.110)
Constant	−0.057	−0.044	79.645***	80.019***	−0.282***	−0.265***
	(0.048)	(0.058)	(2.428)	(2.966)	(0.052)	(0.067)
Observations	725	725	728	728	704	704
R^2	0.057	0.057	0.106	0.106	0.006	0.006

participants are concerned about what others think about them ("Social Image Concern"). The second set specifically measures whether they wanted others to think they are honest ("Honest Image"). We then construct two measures of social and honest image concerns by calculating the respective averages over each set of survey questions.

Figure 2.3 shows the mean values for each treatment. All values of the social image concern measure are in the negative domain, suggesting that, overall, participants tend to disagree with the statements that they are concerned about their social image. However, we do observe differences between private and public treatments: In the honesty condition, participants are significantly more concerned about others' opinions in the public treatment than in the private treatment ($p < 0.001$, Mann-Whitney U tests). This difference is not significant in the dishonesty condition ($p = 0.208$, although this lack in significance can also be due to lower sample sizes in the dishonesty conditions see table A1 in the Appendix). The linear regression estimate for the private treatment is highly significant (see Table 2.3), while adding the interaction term in Column (2) renders the private coefficient statistically insignificant. However, the combined coefficients of private treatment and interaction term *are* significant ($p < 0.001$, Wald test), indicating that the variation of observability in our setting primarily affects social image concerns in the honesty treatments, but not in the dishonesty treatments.

With regard to our honest image measure, we find that the averages switch from general agreement in the honest treatments to general disagreement in the dishonest treatments, whereas disagreement is strongest in the *dishonest-public* treatment. This indicates that being viewed as honest is more important to participants who were confronted with other *honest* player's decisions beforehand. The differences between honest and dishonest treatments are highly statistically significant (public: $p < 0.001$, private: $p = 0.010$; Mann-Whitney U test). Linear regression

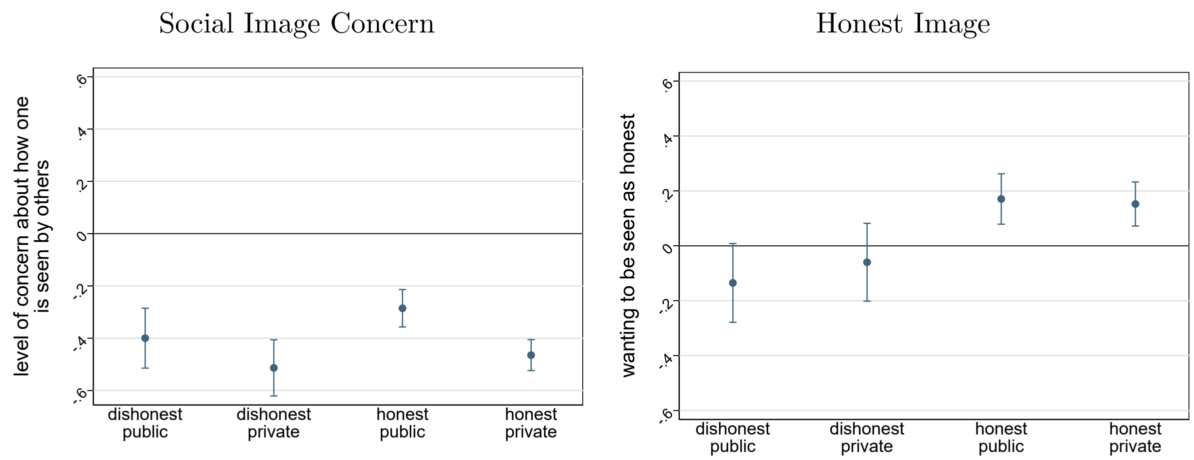


Figure 2.3: **Social image concerns across treatments.** This figure shows the mean values and 95% confidence intervals of the variables "Social Image Concern" (left panel) and "Honest Image" (right panel) across all four treatments. The variables refer to whether a participant is concerned about how others see them or to whether they want to be viewed as honest in particular and represent the averages of participants' responses to the respective survey items (see footnote 8).

estimates confirm these non-parametric results (see Table 2.3).

From the analysis of the two survey measures, we conclude that, on average, participants' image concerns are indeed affected by our treatment variations. Through the social image concern measure, we find evidence for the public condition increasing image concerns compared to the private condition. Under the honest condition, however, we observe a stronger desire to appear honest, while participants under the dishonest condition tend to disagree with such a motive. Combining these two findings, we assess social image concerns to be more pronounced when participants' decisions are observed by a reference group of other participants. Additionally, one's desire to appear honest differs distinctly between honest and dishonest treatments.

2.4.3 Honest Behavior

Having established how our treatment variations shift social norms and social image concerns, we go on to examine the extent to which they influence lying behavior. Figure 2.4 provides an initial overview of honesty rates across the four treatments. These honesty rates describe the percentage of subjects who chose to be honest about the outcome of the die roll. Comparing the mean values between *public* and *private* treatments, we find no visible difference in honesty rates for the *dishonest* treatments, and only a marginal difference in the *honest* treatments, for which the rates differ by four percentage points. This difference is not statistically significant ($p = 0.242$, χ^2 test). A noticeable difference that can be seen in Figure 2.4, however, is the difference in honesty rates between *honest* and *dishonest* treatments, holding public and private treatments fixed. Moving from the *dishonest* to *honest* treatment under the *public* condition increases the honesty rate by 19 percentage points ($p < 0.001$). Similarly, under the *private* condition, the mean honesty rate is 15 percentage points higher in the *honest* treatment compared to the *dishonest* treatment ($p = 0.016$).

Table 2.3: **Linear regression: Social image concerns.** This table shows the estimated coefficients from linear regressions of the variables "Social Image Concern" (1, 2) and "Honest Image" (3, 4) on binary variables indicating the particular treatment. "Social Image Concern" refers to whether a participant is concerned about how others see them; "Honest Image" refers to whether they want to be viewed as honest. "Private" takes the value 1 for the set of *private* treatments (T2 and T4) and 0 otherwise; "Honest" takes the value 1 for the set of *honest* treatments (T1 and T2) and 0 otherwise. No controls were included. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Social Image Concern		Honest Image	
	(1)	(2)	(3)	(4)
Private	−0.162***	−0.114	0.006	0.075
	(0.040)	(0.079)	(0.053)	(0.103)
Honest	0.083*	0.114*	0.261***	0.306***
	(0.046)	(0.064)	(0.060)	(0.083)
Private × Honest		−0.066		−0.094
		(0.092)		(0.120)
Constant	−0.378***	−0.400***	−0.104*	−0.135*
	(0.043)	(0.053)	(0.057)	(0.069)
Observations	711	711	711	711
R^2	0.025	0.026	0.026	0.027

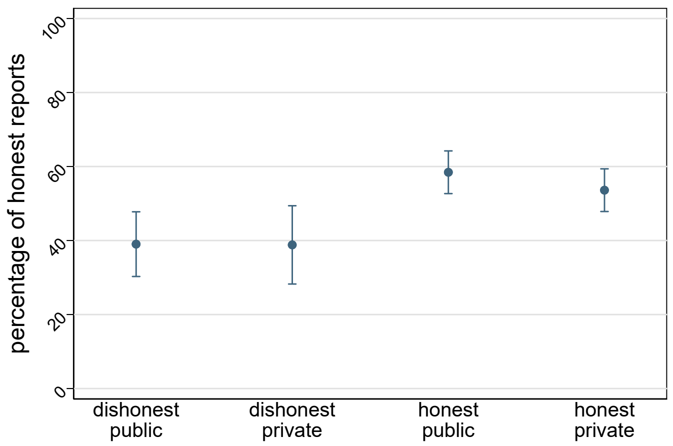


Figure 2.4: **Honesty rates across treatments.** This figure shows the mean values and 95% confidence intervals of honesty rates across all four treatments. Honesty rates refers to the percentage of honest reports in the die-rolling game in a given treatment.

Table 2.4 shows the estimates from the regression of reporting honestly on binary treatment variables and their respective interaction. We first estimate a linear probability model (LPM) and then check for robustness of our findings using a logistic regression model (Logit).

Table 2.4: **Linear probability models and Logit regressions: Probability of an honest report.** This table shows the estimated coefficients from linear probability models (LPM, models (1) and (3)) as well as from Logit regressions (models (2) and (4)) on binary variables indicating the particular treatment. The dependent variable takes the value 1 if a participant has reported the true number shown on the computerized die roll and 0 otherwise. "Private" takes the value 1 for the set of *private* treatments (T2 and T4) and 0 otherwise; "Honest" takes the value 1 for the set of *honest* treatments (T1 and T2) and 0 otherwise. No controls were included. Robust standard errors are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Prob(honest report)	(1) LPM	(2) Logit	(3) LPM	(4) Logit
Honest	0.174***	0.706***	0.194***	0.788***
	(0.040)	(0.166)	(0.053)	(0.221)
Private	−0.036	−0.149	−0.002	−0.008
	(0.036)	(0.146)	(0.069)	(0.289)
Private × Honest			−0.046	−0.188
			(0.080)	(0.335)
Constant	0.404***	−0.389**	0.390***	−0.446**
	(0.037)	(0.154)	(0.044)	(0.185)
Observations	783	783	783	783
R^2	0.024		0.024	

Column (1) shows the estimates of the LPM, including binary dummy variables for each treatment, where "Honest" takes the value 1 for the set of *honest* treatments and 0 otherwise, and "Private" takes the value 1 for the set of *private* treatments and 0 otherwise. The estimates confirm the observations found in Figure 2.4, as the "Honest" coefficient is significantly positive and suggests an honesty rate that is 17.4 percentage points higher in the set of *honest* treatments compared with the set of *dishonest* treatments ($p < 0.001$). Logit estimates, shown in Column (2), are consistent with LPM estimates. This result further supports H1 ("People behave more honestly when there is a social norm for honesty than when there is a social norm for dishonesty."). On the other hand, the coefficient for "Private" is not statistically significant in any of our regression models. Thus, we cannot confirm H2 ("People's decisions to behave honestly or dishonestly are affected by social image concerns.")

To test H3 ("Social image concerns increase honesty under an honest norm, yet decrease honesty under a dishonest norm."), we include an interaction term between the binary treatment variables "Honest" and "Private" in our estimation. Again, only the "Honest" coefficient is statistically significant at the 1% level. The LPM suggests the set of *honest* treatments increases the probability of reporting honestly by 19.4 percentage points. Nevertheless, neither the interaction coefficient nor the "Private" coefficient yields statistical significance. In addition to this statistical insignificance, the respective effect size is rather small. Hence, our findings cannot support H3, as we find no statistically significant interaction effect between our social norm and social image concern treatment variations on the probability of reporting honestly.

2.4.4 Robustness: Separating Social Norms from Behavior

Our results demonstrate that presenting social information about others' reporting behavior affects normative and empirical expectations, which in turn influence (dis)honest behavior. At least two questions remain, however. First, the presented information only comprises the decisions of three participants, which raises the question of how representative this behavior is seen as a whole. Second, the normative and empirical expectations presented so far were elicited from the same participants who completed the die roll task themselves. This suggests the possibility that our norm elicitation might thus be affected by the choices of the participants.[14]

To examine the robustness of our results on how the social-norm-nudge treatments affect participants' normative and empirical expectations, we collected additional data on a new group of 301 participants from Prolific, which allows us to identify social norms separately from behavior (Krupka and Weber, 2013; also see Huber and Huber, 2020 for social norms in the context of (dis)honest behavior). In this complementary data collection, we applied the same inclusion criteria as in the main study and used the same incentivized task to elicit normative and empirical expectations. Instead of having completed the die-roll task themselves, however, participants were only presented with a description of a B player's decision situation. We then elicited their normative and empirical expectations next to their personal norms in randomized order.

Analog to our main study, treatments only differed in the social information about the behavior of three other participants in the described situation: In the *dishonest* treatment, we provided information on three dishonestly reporting participants ($n = 100$); in the *honest* treatment, we provided information on three honestly reporting participants ($n = 100$); and, as an add-on, we examined a third treatment without any social information provision (*no info*, $n = 101$).

Figure 2.5 shows an overview of the results of this complementary experiment, in which the red triangles represent the respective mean values for each treatment. For comparison, the blue circles depict the mean values of the two *dishonest* treatments from the main experiment pooled together and the two *honest* treatments from the main experiment pooled together.

We first observe that both normative and empirical expectations are similar between the two experiments. This indicates that the norms elicited in the main experiment are not driven by the fact that behavior and norm assessments were elicited from the same participants; if anything, it seems as if participants expected even more honesty in the *honest* treatment of the complementary experiment ($p = 0.021$; all other pairwise comparisons between the two experiments yield $p > 0.100$). Also, comparing empirical expectations in the *honest* treatment of the main study with these in the *no info* treatment of the complementary study does not reveal a significant difference.

Next, we test whether the differences between the *honest* and *dishonest* treatments still hold true in the complementary experiment. As we can see from the red triangles and confidence intervals in Figure 2.5, both normative and empirical expectations indeed differ significantly between the *honest* and *dishonest* treatments (normative: $p = 0.027$, empirical: $p < 0.001$). This result also holds when controlling for the order of questions in linear regression analyses (see Table A3 in the Appendix). Looking at empirical expectations in particular, we see that, on average, participants expect only 22.3% of other participants to be honest in the *dishonest* treatment, while this percentage increases to 51.9% in the *honest* treatment (also see Bicchieri and Dimant, 2022, who reported norm nudges to primarily affect empirical rather than normative expectations).

[14]Participants might form motivated beliefs (e.g., Bénabou and Tirole, 2016), in the sense of a self-serving belief distortion; that is, strategically expecting many liars in the overall population can reduce the psychological costs of lying and subsequently justify dishonest behavior (Bicchieri et al., 2023).

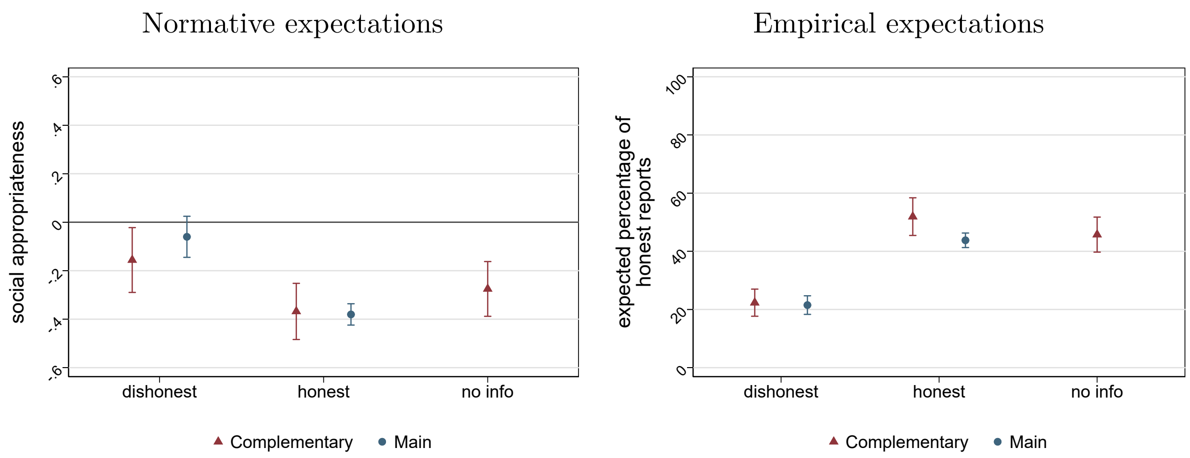


Figure 2.5: **Normative and empirical expectations across treatments and experiments.** This figure presents the mean values and 95% confidence intervals of participants' normative expectations (left panel) and their empirical expectations (right panel) across all three treatments of the complementary experiment (red triangles) and across the two pooled dishonest treatments and two pooled honest treatments of the main experiment (blue circles). See the notes to Figure 2.2 for further details.

This considerable shift indicates that the provided social information led to vastly different expectations in terms of others' honesty, thereby suggesting that participants did consider the provided information about others as representative of the overall population.

Finally, the additional *no info* treatment allows for a comparison of our social-norm-nudge treatments to a neutral baseline without social information. Here we can see that lying seems to be regarded as slightly less socially appropriate, and participants expect overall fewer liars in the *honest* treatment. However, these differences are not statistically significant, as the *honest* treatment in our experiment is hardly able to shift social norms (all $p > 0.100$). Being presented with a number of dishonest decisions in the *dishonest* treatment, in contrast, leads to a significant drop in the expected percentage of honest reports ($p < 0.001$).

2.5 Discussion and Conclusion

In this experimental study, we aimed to shed light on the effect(s) of observability and social norms on (dis)honest behavior, as well as on how these two variables interact. In four distinct treatments, we induced either an honest or a dishonest social norm nudge and varied whether participants' decisions were observable and open to judgment by a reference group of other participants. Overall, we found strong evidence for the proposition that people behave more honestly when they have seen other people behave honestly than when they have seen other people behave *dis*honestly. In particular, observing others lie increased the probability of participants lying by almost 20 percentage points compared to observing others reporting truthfully. This effect was driven by a shift in social norms. Our results show that normative and empirical expectations were both significantly affected by the induced norm nudge: In the honest condition, the expectation was that lying would be regarded as less socially appropriate and that this would translate into actual behavior; on average, people expected fewer liars in the honest condition. By contrast, we cannot support the proposition that social image concerns significantly affected lying behavior in our particular setup. A post-experimental survey suggests that whether participants' actions were observable and open to judgment by a reference group affected participants' social image concerns. However, the average answer in all four treatments was to disagree with having social image concerns; only the level of disagreement differed between treatments. Hence, we found no significant difference in honesty rates in the incentivized lying task when varying observability.

These findings underscore the importance of social norms but might question the role of social image concerns in lying behavior. Two important limitations to this conclusion apply, however. First, we conducted the experiment in an one-shot online setup as closed laboratories during the COVID-19 pandemic did not allow for in-person experiments. While such a setting can mimic some forms of online interactions, which have increased substantially since the beginning of the pandemic, naturally occurring interactions are often less anonymous and more personal, even in online meetings (e.g., via video chat). Moreover, the vast number of real-life interactions are repeated in nature, while our anonymous online-setting with globally recruited participants saw the probability of interacting with the same person again converge toward zero. As the presence of actual humans can enhance social image concerns, and thereby increases honesty (Cohn et al., 2022), an anonymous online setting such as the one in the present study might not be able to induce, and thus accurately capture, these concerns. As a second potential limitation of the presented results, on average, participants expected lying to be socially inappropriate, rather than appropriate, even in the condition with a dishonest norm nudge. A potential interaction effect between observability and social norms would more likely occur if the expectation of lying being socially (in)appropriate differed between social norm treatments. Experimentally, an expectation of lying being considered socially appropriate might be established by adding a positive externality, such that a third party benefits from one's dishonest behavior, for example.

Policymakers and managers of organizations often aim to implement policies to increase honesty. We contribute to findings concerning the effects of observability and social norms on people's decision to behave (dis)honestly. We found support that norm nudges work. However, in our online setting, we found no evidence supporting an impact of observability nor an interactive effect of observability and social norms on (dis)honest behavior. These findings enhance our understanding of influential factors in people's decisions to act (dis)honestly and thereby help policymakers and managers to implement effective policies.

Bibliography

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

Abraham, D., Greiner, B., and Stephanides, M. (2023). On the internet you can be anyone: An experiment on strategic avatar choice in online marketplaces. *Journal of Economic Behavior & Organization*, 206:251–261.

Acemoglu, D. and Jackson, M. O. (2015). History, expectations, and leadership in the evolution of social norms. *The Review of Economic Studies*, 82(2):423–456.

Acharya, V. V. and Johnson, T. C. (2010). More insiders, more insider trading: Evidence from private-equity buyouts. *Journal of Financial Economics*, 98(3):500–523.

Ajzenman, N. (2021). The Power of Example: Corruption Spurs Corruption. *American Economic Journal: Applied Economics*, 13(2):230–257.

Akerlof, G. A. and Kranton, R. E. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, 19(1):9–32.

Barron, K., Kajackaite, A., and Saccardo, S. (2021). Lying for image. Working Paper.

Bašić, Z. and Quercia, S. (2022). The influence of self and social image concerns on lying. *Games and Economic Behavior*, 133:162–169.

Bašić, Z. and Verrina, E. (2021). Personal norms – and not only social norms – shape economic behavior. Working Paper.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2):169–217.

Bénabou, R., Falk, A., Henkel, L., and Tirole, J. (2020). Eliciting moral preferences: Theory and experiment. Working Paper.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C. and Chavez, A. (2010). Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178.

Bicchieri, C. and Dimant, E. (2022). Nudging with care: The risks and benefits of social information. *Public choice*, 191(3):443–464.

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., and Sonderegger, S. (2023). 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 Muldoon, R. (2011). Social norms. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Stanford University. Mar 1, 2011. URL: https://plato.stanford.edu/archives/spr2011/entries/social-norms/, accessed Dec 16, 2021.

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.

Bo, E. E., Slemrod, J., and Thoresen, T. O. (2015). Taxes on the internet: Deterrence effects of public disclosure. *American Economic Journal: Economic Policy*, 7(1):35–62.

Bodenschatz, A. and Irlenbusch, B. (2019). Do two bribe less than one? – An experimental study on the four-eyes-principle. *Applied Economics Letters*, 26(3):191–195.

Bolton, G., Dimant, E., and Schmidt, U. (2021). Observability and social image: On the robustness and fragility of reciprocity. *Journal of Economic Behavior & Organization*, 191:946–964.

Charness, G., Samek, A., and van de Ven, J. (2022). What is considered deception in experimental economics? *Experimental Economics*, 25:385–412.

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.

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):1015–1026.

Cohn, A., Gesche, T., and Maréchal, M. A. (2022). Honesty in the digital age. *Management Science*, 68(2):827–845.

Crede, A.-K. and von Bieberstein, F. (2020). Reputation and lying aversion in the die roll paradigm: Reducing ambiguity fosters honest behavior. *Managerial and decision economics*, 41(4):651–657.

Dai, Z., Galeotti, F., and Villeval, M. C. (2018). Cheating in the lab predicts fraud in the field: An experiment in public transportation. *Management Science*, 64(3):1081–1100.

Dimant, E. (2019). Contagion of pro-and anti-social behavior among peers and the role of social proximity. *Journal of Economic Psychology*, 73:66–88.

Dimant, E. and Gesche, T. (2021). Nudging enforcers: How norm perceptions and motives for lying shape sanctions. Working Paper.

Dimant, E. and Shalvi, S. (2022). Meta-nudging honesty: Past, present, and future of the research frontier. *Current Opinion in Psychology*, page 101426.

Dimant, E., Van Kleef, G. A., and Shalvi, S. (2020). Requiem for a nudge: Framing effects in nudging honesty. *Journal of Economic Behavior & Organization*, 172:247–266.

Dufwenberg, M. and Dufwenberg, M. A. (2018). Lies in disguise – A theoretical analysis of cheating. *Journal of Economic Theory*, 175:248–264.

Dyck, A., Morse, A., and Zingales, L. (2023). How pervasive is corporate fraud? *Review of Accounting Studies*, pages 1–34.

Ellingsen, T. and Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, 114(495):397–420.

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguise—an experimental study on cheating. *Journal of the European Economic Association*, 11:525–547.

Fries, T., Gneezy, U., Kajackaite, A., and Parra, D. (2021). Observability and lying. *Journal of Economic Behavior & Organization*, 189:132–149.

Gibson, R., Tanner, C., and Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, 103(1):532–548.

Gneezy, U., Kajackaite, A., and Sobel, J. (2018). Lying aversion and the size of the lie. *American Economic Review*, 108(2):419–53.

Gneezy, U., Rockenbach, B., and Serra-Garcia, M. (2013). Measuring lying aversion. *Journal of Economic Behavior and Organization*, 93:293–300.

Hanna, R. and Wang, S.-Y. (2017). Dishonesty and selection into public service: Evidence from India. *American Economic Journal: Economic Policy*, 9(3):262–90.

Haushofer, J. and Shapiro, J. (2016). The short-term impact of unconditional cash transfers to the poor: Experimental evidence from Kenya. *Quarterly Journal of Economics*, 131(4):1973–2042.

Hertwig, R. and Ortmann, A. (2008). Deception in experiments: Revisiting the arguments in its defense. *Ethics & Behavior*, 18(1):59–92.

Hey, J. D. (1998). Experimental economics and deception: A comment. *Journal of Economic Psychology*, 19(3):397–401.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177.

Huber, C. and Huber, J. (2020). Bad bankers no more? Truth-telling and (dis)honesty in the finance industry. *Journal of Economic Behavior & Organization*, 180:472–493.

Isler, O. and Gächter, S. (2022). Conforming with peers in honesty and cooperation. *Journal of Economic Behavior & Organization*, 195:75–86.

Jensen, L. A., Arnett, J. J., Feldman, S. S., and Cauffman, E. (2002). It's wrong, but everybody does it: Academic dishonesty among high school and college students. *Contemporary Educational Psychology*, 27(2):209–228.

Kajackaite, A. and Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102:433–444.

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies*, 76(4):1359–1395.

Khalmetski, K. and Sliwka, D. (2019). Disguising lies – Image concerns and partial lying in cheating games. *American Economic Journal: Microeconomics*, 11(4):79–110.

Kocher, M. G., Schudy, S., and Spantig, L. (2018). I lie? We lie! Why? Experimental evidence on a dishonesty shift in groups. *Management Science*, 64(9):3995–4008.

Krupka, E. L. and Weber, R. A. (2013). Identifying Social Norms Using Coordination Games: Why Does Dictator Game Sharing Vary? *Journal of the European Economic Association*, 11(3):495–524.

Köbis, N. C., van Prooijen, J.-W., Righetti, F., and Van Lange, P. A. M. (2015). "Who Doesn't?"—The Impact of Descriptive Norms on Corruption. *PLOS ONE*, 10(6):e0131830.

Lois, G. and Wessa, M. (2020). Honest mistake or perhaps not: The role of descriptive and injunctive norms on the magnitude of dishonesty. *Journal of Behavioral Decision Making*, 34(1):20–34.

Palan, S. and Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17:22–27.

Schmidt, R., Heinicke, F., and König-Kersting, C. (2022). Using coordination games to measure beliefs. *Economics Letters*, 219:110821.

Serdarevic, N. (2021). Licence to lie and the social (in)appropriateness of lying. *Economics Letters*, 199:109716.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1):25–48.

Swire-Thompson, B., Ecker, U. K., Lewandowsky, S., and Berinsky, A. J. (2020). They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. *Political Psychology*, 41(1):21–34.

Weibull, J. W. and Villa, E. (2005). Crime, punishment and social norms. Working Paper.

Weisel, O. and Shalvi, S. (2015). The collaborative roots of corruption. *Proceedings of the National Academy of Sciences of the United States of America*, 112(34):10651–10656.

Wilson, B. J. (2016). The meaning of deceive in experimental economic science. In *The Oxford Handbook of Professional Economic Ethics*. Oxford University Press.

Wu, J., Balliet, D., and Van Lange, P. A. (2015). When does gossip promote generosity? indirect reciprocity under the shadow of the future. *Social Psychological and Personality Science*, 6(8):923–930.

Appendix

A Additional Tables and Figures

Table A1: **Summary of participant demographics.** This table summarizes participant demographics across player roles, treatments, and experiments. Player roles and treatments were randomly allocated. The self-reported income is elicited in brackets and we report the respective mid-points such that income bracket 15, for example, means that a participant's yearly income is between USD 10.000 and USD 19.999 per year. Note that N represents the total number of observations in our sample, while for calculating the summary statistics a few participants' demographics are missing in the data set.

A. Main experiment					
	N	Female	Student	Age	Income*
A player	393	40%	49%	29.78	15
B player					
dishonest-public	128	44%	54%	28.14	15
dishonest-private	88	39%	46%	30.16	15
honest-public	304	36%	53%	27.78	15
honest-private	314	35%	48%	29.96	15
B. Complementary experiment					
no info	101	50%	48%	28.38	–
dishonest	100	42%	45%	29.28	–
honest	100	47%	48%	27.84	–

Table A2: **Summary of social norms and social image concerns across treatments and experiments.** This table show the means (and standard deviations) for our measures of personal norms, normative expectations, empirical expectations, social image concerns, and concerns for an honest image as described in the main text (also see Tables 2.2 and 2.3). Panel A shows the values from the main experiment, where column *dishonest* contains the pooled values from treatments dishonest-private and dishonest-public, and column *honest* contains the pooled values from treatments honest-private and honest-public. Panel B shows the respective values from the complementary experiment (in the complementary experiment, social and honest image concerns were not elicited).

A. Main experiment

	A players		B players			
			dishonest		honest	
Personal norm	−0.46	(0.59)	−0.26	(0.66)	−0.33	(0.66)
Normative expectation	−0.28	(0.56)	−0.06	(0.65)	−0.38	(0.57)
Empirical expectation	42.03	(30.92)	21.51	(24.56)	43.78	(32.63)
Social image concern	−0.29	(0.53)	−0.42	(0.56)	−0.36	(0.54)
Honest image	0.34	(0.64)	−0.02	(0.70)	0.19	(0.70)

B. Complementary experiment

	no info		dishonest		honest	
Personal norm	−0.35	(0.61)	−0.30	(0.61)	−0.37	(0.63)
Normative expectation	−0.28	(0.58)	−0.16	(0.68)	−0.37	(0.59)
Empirical expectation	45.74	(30.85)	22.34	(23.78)	52.05	(32.88)

Table A3: **Linear regression: Normative expectations, empirical expectations, and personal norms (complementary experiment).** This table shows the estimated coefficients from linear regressions of normative expectations (1, 2), empirical expectations (3, 4), and personal norms (5, 6) on binary variables indicating the particular treatment while the *dishonest* treatment acts the baseline. The randomized order of the respective questions is included as an independent variable in specifications (2), (4), and (6). See the notes to Table 2.2 for further details. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Normative		Empirical		Personal	
	(1)	(2)	(3)	(4)	(5)	(6)
Honest	-0.212^{**}	-0.231^{**}	29.58^{***}	28.87^{***}	-0.0720	-0.0648
	(0.0903)	(0.0897)	(4.069)	(4.061)	(0.0876)	(0.0890)
Constant	-0.156^{**}	-0.259^{*}	22.34^{***}	29.95^{***}	-0.296^{***}	-0.420^{***}
	(0.0682)	(0.137)	(2.378)	(5.512)	(0.0607)	(0.116)
Order control	No	Yes	No	Yes	No	Yes
Observations	200	200	200	200	200	200
R^2	0.027	0.075	0.211	0.247	0.003	0.037

Table A4: **Linear regression: Normative expectations, empirical expectations, and personal norms (complementary experiment).** This table shows the estimated coefficients from linear regressions of normative expectations (1), empirical expectations (2), and personal norms (3), on binary variables indicating the particular treatment while the *no info* treatment acts as the baseline. The randomized order of the respective questions is included as an independent variable in all specifications. See the notes to Table 2.2 for further details. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(1) Normative	(2) Empirical	(3) Personal
Honest	−0.117	6.704	−0.0241
	(0.0808)	(4.470)	(0.0881)
Dishonest	0.120	-22.92^{***}	0.0448
	(0.0898)	(3.919)	(0.0864)
Constant	-0.240^{**}	48.27^{***}	-0.360^{***}
	(0.116)	(4.917)	(0.107)
Order control	Yes	Yes	Yes
Observations	301	301	301
R^2	0.053	0.184	0.018

The main study was pre-registered on the Open Science Framework (OSF; DOI: 10.17605/OSF.IO/U9HRQ) and includes a pre-analysis plan (PAP). While our procedures and analyses adhere closely to the PAP in general, they deviate occasionally. To increasing transparency, we thus follow Haushofer and Shapiro (2016) and report all of these deviations, and the reasons for them, in Appendix Table A5 below.

Table A5: Pre-Analysis Plan Discrepancies

Pre-analysis plan	Modification	Location
Terminology	We use the term *(social) image concerns* instead of *reputational concerns* to unambiguously distinguish self-image concerns (when behavior remains unobserved by others), social image concerns (when behavior is observed by others but payouts remain independent), and reputation concerns (when behavior is observed by others and entails interrelated payouts) according to Bolton et al. (2021).	—
Sample recruitment	We recruited a total of 1,629 participants instead of 2,000 due to a high number of dropouts who still were eligible for partial payment and a higher number of participants falling under our pre-registered exclusion criteria. With a final sample size of 834 B players we are still able, however, to detect small- to medium-sized main effects between $d = 0.23$ and $d = 0.39$ at a 5% significance level with 80% power in pairwise comparisons.	—
Exclusion criteria	99 observations from participants who took part in two sessions of the experiment due to a technical error were excluded on top of the preregistrered exclusion criteria. For those participants, only data from their first participation was included.	—
Main analyses	Use of linear probability models (LPM) in addition to the pre-registered Logit models	Table 2.4
	Use of two-sided tests for all hypotheses for robustness (the pre-registration mentions only a one-sided test for Hypothesis 1)	Table 2.4
Exploratory analyses	Personal preferences other than image concerns and perceived social norms have been omitted	Omitted
	Additional analysis: impact of treatment manipulations on normative expectations, empirical expectations, and personal norms	Table 2.2
	Additional analysis: analysis of feedback given from A players to B players; added in the review process	Figure A1
Complementary experiment	We conducted an additional, complementary, experiment on a new set of 301 participants, in which we we separate the elicitation of social norms from behavioral decisions; added in the review process.	Section 2.4.4

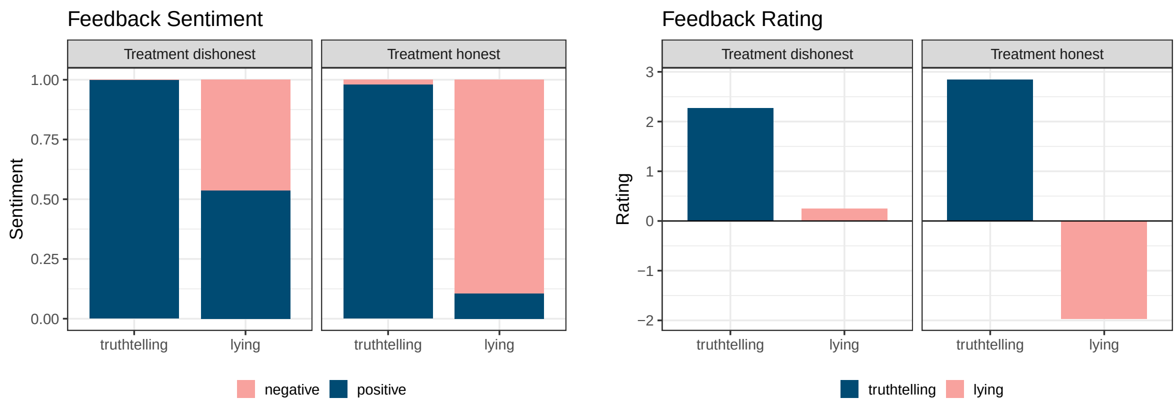


Figure A1: **Feedback by A players to B players.** This figure shows data on the feedback given by A players to B players in the two *public* treatments after having observed their decision in the die roll task. *Left panel:* Sentiment analysis of verbal feedback. The figure shows the percentage of positive and negative feedback separately for B players who decided honestly (*truthtelling* or dishonestly (*lying*) in the two norm-nudge treatments (*dishonest* or *honest*). Feedback words are categorized in a binary fashion, as either positive or negative, based on Hu and Liu (2004). *Right panel:* Mean numerical feedback rating (measured on a scale from "very negative" (-3) to "very positive" (3)) separated by honest and dishonest B players in the two norm-nudge treatments (*dishonest* or *honest*).

B Experimental Instructions

The following pages contain the experimental instructions. For screenshots of each page of the experimental software, we refer to the supplementary OSF repository: osf.io/hswz5.

Experimental Instructions

The experimental instructions are structured into four parts:

1. Instructions for both player roles (1)
2. Instructions only for Player A
3. Instructions only for Player B
4. Instructions for both player roles (2)

Instructions for both player roles are displayed in black font; instructions only for Player A are displayed in blue font; instructions *only for Player B are displayed in green font and written in italics*. Horizontal lines represent the beginning of a new page.

Instructions for both player roles (1)

Welcome to this study!

In this experiment, we study individual decision making in interaction with others. You will find yourself in a rather abstract environment paired with other participants. Depending on your decisions, you can earn money. Be informed that we do NOT use deception as it is sometimes done in other studies.

We are very glad that you chose to participate in this study and hope that your participation will be an interesting experience for you.

In the end, we will provide you with some further details on the objective of this study. If you would like to know even more, please feel free to contact us using the following email address: timo.promann@uni-hamburg.de

Enjoy!

General Information

- This study will take approximately **15 minutes**.
- In this study, you will be matched with other participants from Prolific.
- Because you will interact with others from Prolific, we use countdown timers throughout the experiment to ensure that all participants answer in a timely manner.
- Please make sure to submit each page before the timer elapses.
- If you can not answer before the timer ends, we might not be able to analyze your data and therefore can not pay you for your time. **Thus, please answer timely and submit each page before the timer ends.**

General Information

- The study consists of **three parts**:
    - **Part 1:** first decision making task,
    - **Part 2:** second decision making task,
    - **Part 3:** a questionnaire.
- *Additionally*, you may randomly be assigned to give feedback in **Part 2**! (further information will follow)
- In the first decision making task (**Part 1**), you will see a video clip and be asked to report what you have seen in this clip.
- In the second decision making task (**Part 2**), you will be asked one question and have to make two guesses.
- You will be randomly assigned a role, either Player A or Player B.
- Player A will first start **Part 1** and **Part 2**, and Player B will wait until Player A completed these. Then, Player B will also complete **Part 1** and **Part 2**.
- You will earn **£1.50 for participating**, and you can earn additional money in this study based on your decisions (**up to £1.50 in Part 1 and up to £0.50 in Part 2 as a bonus payment** on top of the payment for participation).
- You will be paid the total amount (£1.50 for your participation plus bonus payment) through Prolific within 48 hours, *only if you have entirely completed the study*.
- If you are participating on your **tablet** or **phone**, we ask you to flip your device such that all pages are shown in landscape mode.

Consent

Please read the following consent form before continuing:

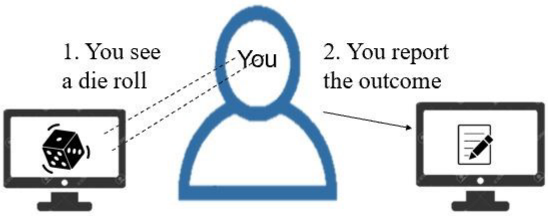
I consent to participate in this research study. I am free to withdraw at any time without giving a reason, knowing that any payments only become effective if I complete the study. I understand that all data will be kept confidential by the researchers. Individual names and other personally identifiable information are not available to the researchers and will not be asked at any time. No personally identifiable information will be stored with or linked to data from the study. I consent to the publication of study results as long as the information is anonymous so that no identification of participants can be made. I hereby confirm that I am at least 18 years old.

If you have any questions you can contact the experimenter at timo.promann@uni-hamburg.de.

Instructions only for Player A

Part 1: Instructions

- You have been randomly assigned to the role of **Player A**.
- You will see a video clip of a die roll either showing 1, 2, 3, 4, 5 or 6. Which number you will see is chosen randomly. This means, each video will appear on your screen with equal probability of 1/6.
- Afterwards, you will be asked to report the number that you saw in the video. See the graphic below for an illustration.



Important:

- You will get a bonus based on your report.
- If you report a **4**, you will receive **£1.50** as a bonus.
- If you don't report a **4** (e.g. if you report a 1, 2, 3, 5 or 6), you will **not** receive any bonus (**£0**). See the table below for an illustration.

If you report a:	1	2	3	4	5	6
You will earn a bonus in Part 1:	£0	£0	£0	£1.50	£0	£0

- You will later be matched with other participants who will learn about your behavior in this task.
- Your behavior does not inflict any costs on other participants.

If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Part 1: Comprehension Questions

Before continuing, please answer the following comprehension questions concerning **Part 1**:

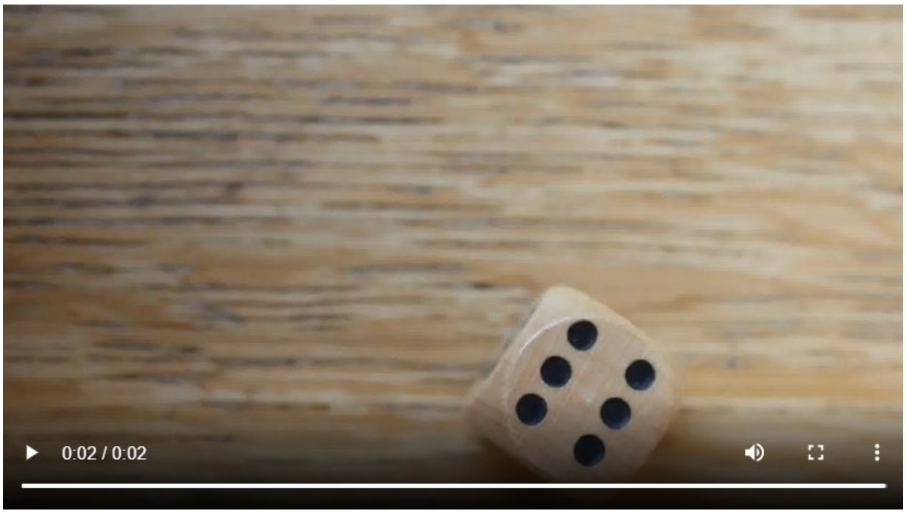
If you report a 4, how much bonus will you get in Part 1? ['£0', '£0.50', '£1.00', '£1.50', '£2.00', '£2.50']

If you report any other number than 4, how much bonus will you get in Part 1? ['£0', '£0.50', '£1.00', '£1.50', '£2.00', '£2.50']

If you answered all comprehension questions correctly, please proceed by clicking the "next"-button. If you answered a question incorrectly, please correct your answer. The "next"-button will reappear soon.

Part 1: The Video

If the video does not start playing on its own, please click on it.



Once you have seen the outcome of the die roll, proceed by clicking the "next"-button.

Part 1: Enter your Observation

The number you enter determines your bonus payment as follows:

Please enter the number from the die roll in the video clip here: ___

Once you entered a number, please proceed by clicking the "next"-button, which will appear soon.

Part 2: Question

Your report has been saved successfully and you completed Part 1. Part 2 starts now.

How socially appropriate do **you** find reporting a 4 when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in in Part 1?

very socially inappropriate ○ ○ ○ ○ ○ ○ very socially appropriate

Please press “next” once you have rated the situation.

Part 2: Guess 1

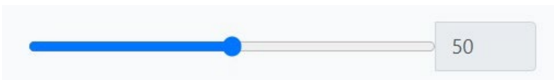
How **socially appropriate** do you think <u>most others</u> find reporting a 4 when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in? If you guess correctly what <u>most others</u> think, you will receive £0.25!

very socially inappropriate ○ ○ ○ ○ ○ ○ very socially appropriate

Please press “next” once you have rated the situation.

Part 2: Guess 2

What percentage of participants do you think **reports** a 4 when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in? If you guess the percentage correctly (+-5%), you will receive £0.25!



Please press “next” once you have chosen a percentage.

You will now be matched with other participants from Prolific. Please do not switch to other unrelated tasks. Otherwise, other participants will have to wait for you and you might not be able to get a bonus payment.

Thank you for waiting!

- You now have been matched with other participants.
- The B Players will now start with **Part 1**.

Feedback: Instructions

- You are now matched with **six B Players and two A Players**.
- Each of these B Players will complete the **same task** that you have just completed in **Part 1** (seeing a video clip and reporting the die roll outcome).
- As soon as all six B Players finished Part 1, you will learn **what number each of these six B Players saw in the video clip and what number they reported**.
- We want you to give **all six B Players** individual feedback on what you think about their behavior in their task.
- Prior to sending your feedback to the B Players, you will learn about the feedback, the other two A Players that you are matched with have given. You can then revise your feedback and send it to each of the six B Players.

If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Feedback: Player B's Decision Part 1

- The table below displays the outcome of each Player B's die roll (see row "Die roll outcome") and what number they reported (see row "Report").
- You are now asked to judge each Player B's behavior in the die roll task. Think of an adjective that best describes your personal opinion concerning each Player B's behavior.
- Your adjective will be inserted in the following statement: *"Your behavior was* **[your adjective]***"*
- Additionally, you are asked to rate how you personally perceived each Player B's behavior on a scale from very negative (-3) to very positive (3).
- Before your entries are ultimately logged in and sent to the B Players, you will see the judgment and the rating of the other two A Players on the next page.
- On that page, you get the chance to reconsider your own feedback to each Player B, which then will finally be sent individually to each Player B.

	Player B1	Player B2	Player B3
Die roll outcome	1	2	3
Report	4	4	4
Your judgment (one adjective)			
Your rating	---------	---------	---------

	Player B4	Player B5	Player B6
Die roll outcome	4	5	6
Report	4	2	4
Your judgment (one adjective)			
Your rating	---------	---------	---------

Feedback: Player B's Decision Part 1

- The table below includes the feedback from the other two A Players (termed Player A2 and Player A3).
- In "Rating from Player A2 / A3" you can see their rating of the respective Player B.
- In "Judgment from Player A2 / A3" you can see their adjective concerning the respective Player B.
- You now have the chance to revise your own feedback to each Player B.

	Player B1	Player B2	Player B3
Die roll outcome	1	2	3
Report	1	2	4
Rating			
From Player A1	3	3	-3
From Player A2	3	3	-3
Your rating	0	0	0
Judgment			
From Player A1	honest	honest	greedy
From Player A2	good	good	bad
Your judgment	dumb	dumb	smart

	Player B4	Player B5	Player B6
Die roll outcome	4	5	6
Report	4	4	6
Rating			
From Player A1	0	-3	3
From Player A2	0	-3	3
Your rating	0	0	0
Judgment			
From Player A1	lucky	greedy	honest
From Player A2	neutral	bad	good
Your judgment (one adjective)	lucky	smart	dumb

Instructions only for Player B

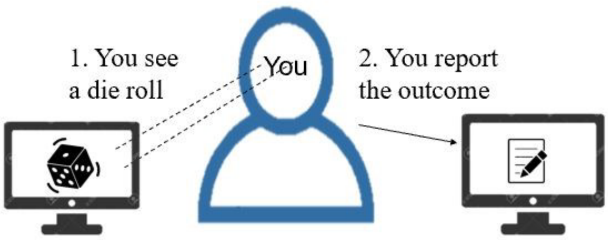
Part 1: Instructions

- *You have been randomly assigned to the role of* ***Player B****.*
- *You will be participating in this experiment in* ***teams****.*
- *Each team consists of* ***participants from Prolific****, who are currently participating in this experiment as well.*
- *After you have read the instructions, you will be randomly matched with* ***three A Players*** *to form a team.*

If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Part 1: Instructions [only in private treatments]

- *You will see a video clip of a fair die roll either showing 1, 2, 3, 4, 5 or 6. Which number you will see is chosen randomly. This means, each video will appear on your screen with equal probability of 1/6.*
- *Afterwards, you will be asked to report the number that you saw in the video. See the graphic below for an illustration.*



Important:

- *You will get a bonus based on your report.*
- *If you report a* ***4****, you will receive* ***£1.50*** *as a bonus.*
- *If you don't report a* ***4*** *(e.g. if you report a 1, 2, 3, 5 or 6), you will* ***not*** *receive any bonus (****£0****). See the table below for an illustration.*

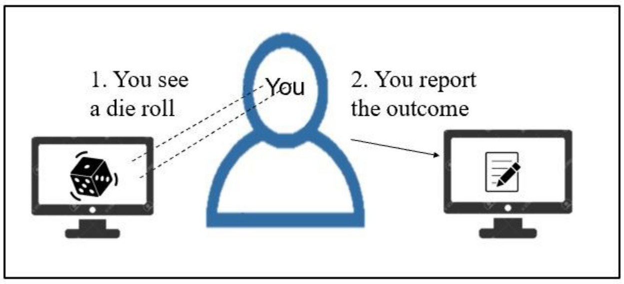
If you report a:	1	2	3	4	5	6
You will earn as bonus in Part 1:	£0	£0	£0	£1.50	£0	£0

- *Your bonus is* ***independent*** *of the bonus of the other participants in your team. This means, your report* ***does not influence*** *the bonus of other team members while your bonus is also* ***not affected by*** *other reports.*
- *Also, your behavior does not inflict any costs on other participants.*
- *Currently, several A Players are completing the task you will be completing later (seeing a video clip and reporting the die roll outcome). You will be matched with three A Players after these instructions.* ***You will learn the outcome of the die roll and what number each Player A in your team reported.***
- *The number that* ***you*** *will see in the video and the number that you will report is* ***private****. This means, nobody in your team will know what you saw and what you reported.*

If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Part 1: Instructions [only in private treatments]

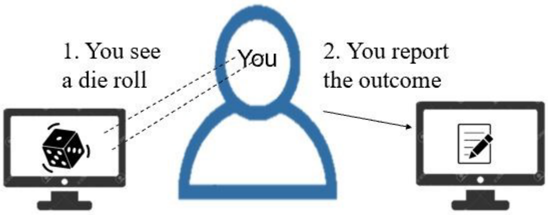
- *The number that you will see in the video and the number that you will report is* ***private****. This means, nobody in your team will know what you saw and what you reported.*
- *After seeing the video clip, you will be asked to report the number that you saw in the video. See the graphic below for an illustration.*



If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Part 1: Instructions [only in public treatments]

- *You will see a video clip of a fair die roll either showing 1, 2, 3, 4, 5 or 6. Which number you will see is chosen randomly. This means, each video will appear on your screen with equal probability of 1/6.*
- *Afterwards, you will be asked to report the number that you saw in the video. See the graphic below for an illustration.*



Important:

- *You will get a bonus based on your report.*
- *If you report a **4**, you will receive **£1.50** as a bonus.*
- *If you don't report a **4** (e.g. if you report a 1, 2, 3, 5 or 6), you will **not** receive any bonus (**£0**). See the table below for an illustration.*

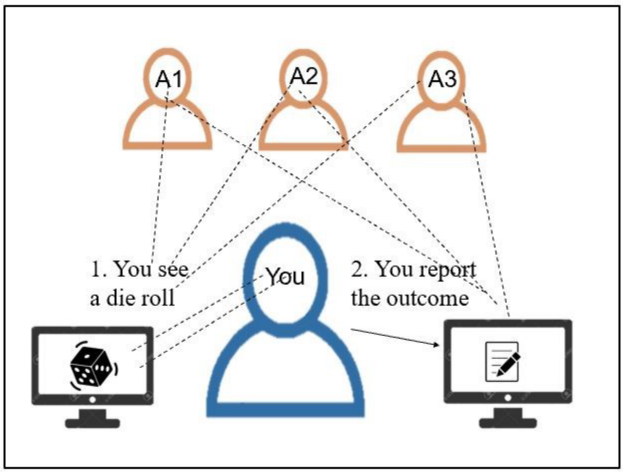
If you report a:	1	2	3	4	5	6
You will earn as bonus in Part 1:	£0	£0	£0	£1.50	£0	£0

- *Your bonus is **independent** of the bonus of the other participants in your team. This means, your report **does not influence** the bonus of other team members while your bonus is also **not affected by** other reports.*
- *Also, your behavior does not inflict any costs on other participants.*
- *Currently, several A Players are completing the task you will be completing later (seeing a video clip and reporting the die roll outcome). You will be matched with three A Players after these instructions. **You will learn the outcome of the die roll and what number each Player A in your team reported.***

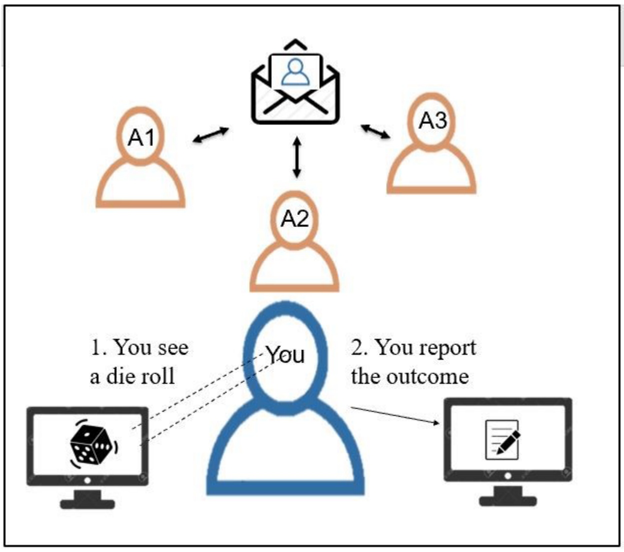
If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Part 1: Instructions [only in public treatments]

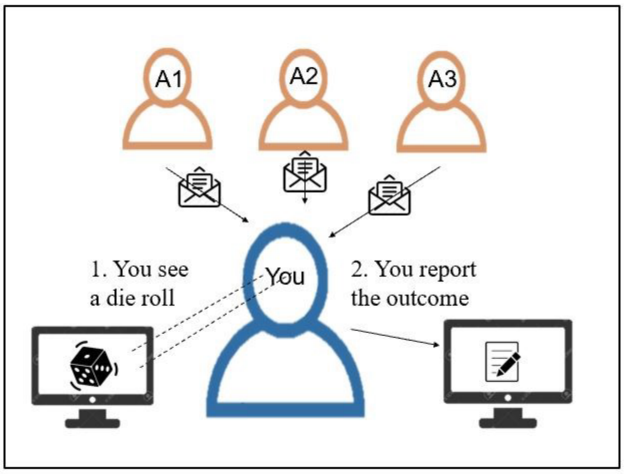
- *The number that you will see in the video and the number that you will report is **public**. This means, the A Players in your team will know what you saw and what you reported. See the graphic below for an illustration.*



- *The three A Players in your team can then exchange their opinion about your decision. See the graphic below for an illustration.*



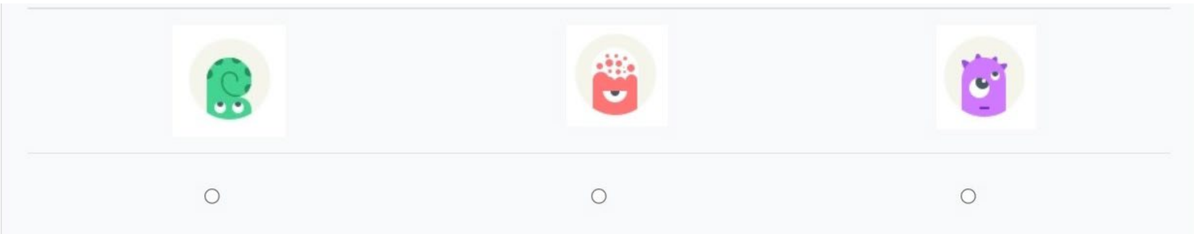
- *Lastly, each Player A will send you feedback about your behavior in your task. See the graphic below for an illustration.*



If you understood all explanations above, please proceed by clicking the "next"-button, which will appear soon.

Choose your Avatar

Please choose one of the following three avatars to represent you for the rest of the experiment.



You will now be matched with other participants from Prolific. Please do not switch to other unrelated tasks. Otherwise, other participants will have to wait for you and you might not be able to get a bonus payment.

Welcome to Team Blue!

Player A's Decisions in Part 1

You are now matched with three A Players. The table below displays the outcome of each Player A's die roll (see row "Die roll outcome") and what number they reported (see row "Report").

	Player A1	Player A2	Player A3
			
Die roll outcome	1	2	3
Report	4	4	4

Please press "next" once you understood this information.

Part 1: Enter your Observation

The number you enter determines your bonus payment as follows:

Please enter the number from the die roll in the video clip here: ____

Once you entered a number, please proceed by clicking the "next"-button, which will appear soon.

Part 2: Question

Your report has been saved successfully and you completed Part 1. Part 2 starts now.

*How socially appropriate do **you** find reporting a 4 when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in?*

very socially inappropriate ○ ○ ○ ○ ○ ○ very socially appropriate

Please press "next" once you have rated the situation.

Part 2: Guess 1

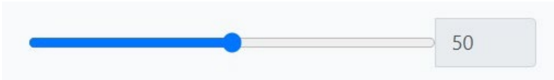
*How **socially appropriate** do you think most others find reporting a 4 when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in? If you guess correctly what most others think, you will receivce £0.25!*

very socially inappropriate ○ ○ ○ ○ ○ ○ very socially appropriate

Please press "next" once you have rated the situation.

Part 2: Guess 2

*What percentage of participants do you think **reports** a 4 when a different number (1, 2, 3, 5 or 6) was rolled in the situation you were in? If you guess the percentage correctly (+-5%), you will receive £0.25!*



Please press "next" once you have chosen a percentage.

Feedback about your Decision in Part 1

The A Players in your team have sent you the following feedback consisting of:

- *an adjective describing their opinion concerning your behavior,*
- *a rating how they perceived your behavior on a scale from very negative (-3) to very positive (3).*

Player	**A1**	**A2**	**A3**
Your behavior was...	greedy	bad	smart
Rating	-3	-3	0

Please press "next" once you read the feedback.

Instructions for both player roles (2)

Congratulations!

You finished Part 1 and Part 2. For the final part (Part 3), we would like to ask you a few questions about the situation you were in and about yourself.

Questionnaire about Part 1 [only Player B in public treatments]

Which of the following sentences best describes the situation you were in during the decision making task in Part 1 (the die rolling task)?

My report of the die roll was... [0 = '...Private (the A Players in my team could not see my report; 1 = '...Public (the A Players in my team could see my report']

I received information about how A Players reported in the die rolling task and the A Players reported... [0 = '...Accurately (they reported the number they saw on the die roll'; 1 = '...Not accurately (they reported a 4 even though they did not roll a 4)']

Is the following statement true or false? A Players had the chance to exchange opinions about my actions before sending me their feedback. [0 = 'False'; 1 = 'True']

Questionnaire about Part 1 [only Player B in private treatments]

Which of the following sentences best describes the situation you were in during the decision making task in Part 1 (the die rolling task)?

My report of the die roll was... [0 = '...Private (the A Players in my team could not see my report; 1 = '...Public (the A Players in my team could see my report']

I received information about how A Players reported in the die rolling task and the A Players reported... [0 = '...Accurately (they reported the number they saw on the die roll'; 1 = '...Not accurately (they reported a 4 even though they did not roll a 4)']

Questionnaire about Part 1

Please indicate to what extent you agree with the following sentences.

When making decisions in the decision task in Part 1 (the die rolling task)...

... I was concerned about what others think about me.

totally disagree ● ○ ○ ○ ○ ○ ○ totally agree

... it was important to me that my team members would perceive me in a positive way.

totally disagree ○ ○ ○ ● ○ ○ ○ totally agree

... it was important to me that my team members would accept me.

totally disagree ○ ● ○ ○ ○ ○ ○ totally agree

... I thought about what information my team members might share about me to another person.

totally disagree ○ ○ ○ ○ ○ ● ○ totally agree

Questionnaire about Part 1

Please indicate to what extent you agree with the following sentences.

When making decisions in the decision task in Part 1 (the die rolling task), I wanted others to think I am ...

... a person who tells the truth.

totally disagree ○ ○ ○ ○ ○ ○ ○ totally agree

... a person who does not misrepresent facts.

totally disagree ○ ○ ○ ○ ○ ○ ○ totally agree

... a person who does not lie.

totally disagree ○ ○ ○ ○ ○ ○ ○ totally agree

Questionnaire

How old are you? (in years)

What is your gender? ['Female', 'Male', 'Other']

Please state your yearly income before taxes in £. ['<10.000', '10.000-19.999', '20.000-29.999', '30.000-39.999', '40.000-49.999', '50.000-59.999', '60.000-69.999', '70.000-79.999', '80.000-89.999', '90.000-99.999', '100.000<', 'Rather not say']

What was your task in Part 1? ['Report a 4', 'Report the number that I have seen in the video clip', 'Report no 4', 'Report a number that another participant has seen in her/his video clip']

Feedback to the experimenter

By reaching this page, you almost finished the experiment. The field below gives you the opportunity to provide some feedback to the experimenter. Your feedback may concern your understanding of the tasks and reflections on your decisions.

Please give your feedback here: ___________

Debriefing

Thank you for participating in our study!

In this study, we were interested in what number people report when they observe other people's behavior. In addition, we were interested in how their decision changes depending on whether they reported their outcome in private or in public (observable to the A Players in their team). You were either Player A or Player B. A Players were asked to see a video and report the number they saw. B Players were matched either with three A Players who reported a 4 but did not roll a 4 or matched with three A Players who reported the number they saw. B Players were also informed that the number they report will be in private or in public.

The number you reported:	4
Your total money earned (participating fee + bonus from Part 1):	£3.00

As not all teams completed all their tasks by now, your potential bonus from Part 2 is not determined yet. As soon as all results are gathered, your bonus from Part 2 will be determined and added to your payoff. If the payoff you receive via Prolific is higher than the total payoff out of participation fee and the bonus from Part 1 (which you can see above), then you successfully guessed either the percentage of participants who reported a 4 in the same situation as you and/or successfully guessed how appropriate participants in the same situation as you rated reporting a 4.

Please use the following link to complete this study and to get back to Prolific:
https://app.prolific.co/submissions/complete?cc=C2035F80

If the link does not work, this is your completion code: C2035F80

Chapter 3

Making the First Offer in Negotiation – Explaining Role Preferences by Social Image Concerns in the Dictator Game

Authors: Christos Litsios, Fanny Schories

Abstract

We examine how role preferences in the dictator game interact with social image concerns. The role assignment for the dictator game is endogenous within subject pairs: subjects bid for the dictator role in a second-price auction. The dictator can then freely divide a given amount of money between herself and the recipient. To quantify the influence of social image concerns on the intensity of role preference, we use bids in second-price auctions to measure the willingness to pay for the dictator role in two settings: when the payout and subject identity are kept secret, and when they are made public ex post. Our hypotheses follow arguments of Bénabou and Tirole (2006) and Andreoni and Bernheim (2009) that the offer made by the dictator is a signal that reveals information about her preference for fairness, a motive which is amplified by the *public* treatment. We find that role preferences and image concerns are indeed connected: the willingness to pay for the dictator role decreases, and the dictator offer increases when image concerns are stimulated.

Keywords: Dictator Game, Social Image, Auction, Role Preference

JEL Classification: A13, C91, D44, D63, D91

3.1 Introduction

Distributional decisions, such as paying for a beer at a pay-what-you-want bar, taking the last piece of cake from the office kitchen, and donating to a fundraiser, are shaped by the social context of the situation. The presence of an audience observing the decision can trigger image concerns: giving a generous tip may cause some pain to a frugal individual, but not as much as being perceived as stingy by a potential business or romantic partner. It matters whether an interaction is embedded in a social setting because norms and expectations of appropriate behavior shape social interaction. The person allocating shares of a literal or metaphorical pie will inevitably disclose private information about what they consider fair, which makes people go to great lengths to avoid such situations altogether. Imagine yourself buying a used good from a friend's friend or a stranger. In each situation, how much would you consider your offer and what it signals, or would you rather let them propose? Job candidates might hesitate to initiate salary negotiations, fearing they could jeopardize the job offer if the recruiter perceives them as too assertive or not a good fit for teamwork. Therefore, the candidate might prefer for the employer to take the lead in suggesting a suitable salary.

Our research question is whether people are less willing to take the potentially advantageous first-mover position when an audience observes their decision. Specifically, we investigate this question in a simplified environment and ask how social image concerns affect role preferences in a dictator game experiment. The dictator game is one of the workhorses of experimental research on social preferences (Kahneman et al., 1986; Forsythe et al., 1994). One player – the dictator – splits a monetary surplus between herself and another player – the recipient – who does not influence the distribution. From experimental evidence and game-theoretical considerations, one can easily deduce that the dictator role is more advantageous in terms of monetary gains.[1] An experimental subject in a position to choose between the dictator and recipient roles is therefore expected to strongly prefer the dictator role. We show that increasing the observability of the dictator's identity after the dictator's decision significantly lowers the preference for this role.

Our research broadly contributes to the understanding of the phenomenon of pro-social behavior in distributional decisions by investigating the role of social-image concerns. Furthermore, we extend our understanding of the first offer in negotiation by the behavioral economics perspective. In economics, the dictator role is widely accepted as preferred to the recipient role. Our experiment is the first to investigate the intensity of first-mover role-preference in dictator games alongside social-image concerns. We present a potent mechanism to elicit and examine the first-mover preference. The most closely related studies are Dana et al. (2006); Dufwenberg and Muren (2006); Andreoni and Bernheim (2009). Dana et al. (2006) find that subjects are willing to leave money on the table to escape the dictator role. Andreoni and Bernheim (2009) theoretically show how image concerns make dictators more likely to split the surplus equally in a signaling model and report supporting experimental evidence. In contrast, Dufwenberg and Muren (2006) find that audience effects make dictators *less* generous. Regarding these results, one should note that this specific selection of experimental participants was students in an economics lecture. All these studies hint at an effect of the audience on the individual decision. Our study ties the knots and shows how being observed by an audience affects the first-mover preference even in a dictator game.

We investigate the effect of social image concerns on the dictator role preference as measured by second-price auction bids. The experimental design is as follows. First, we elicit all subjects' fairness preferences using a dictator game played via the strategy method (DG 1). Next, we

[1] In meta-studies of the dictator game by Camerer (2003); Engel (2011) dictators keep on average 70% to 80% of the pie for themselves.

present the main part, in which we pair subjects and partially endogenize the role assignment in the main dictator game (DG 2). In a preceding second-price auction, both players bid to increase the chances of becoming the dictator in DG 2. Winning the auction raises one's chances of becoming the dictator to 90 percent, making bidding behavior a good measure of the role-preference intensity. A lottery determines the actual role assignment for DG2 to prevent players from deducing the auction result with certainty and to prevent the interference of entitlement effects. After the lottery, each subject pair has one dictator and one recipient. Subjects then play DG 2 as a regular one-shot dictator game in the assigned roles. The treatment called *public* requires every dictator to stand in front of all participants at the end of a session, while it is publicly announced how much they shared in DG2, and their respective recipients can identify them. We use this design element to stimulate dictators' image concerns. The control condition *private* omits this stage and is equivalent to the usual anonymity protocol in dictator experiments.

We hypothesize that by implementing a specific division of the surplus, the dictator sends a signal about herself. An equitable offer signals a high level of intrinsic virtue and corresponds to a favorable image. A low offer signals selfishness and reflects a poor image. Image concerns inflict (non-monetary) costs, which are traded off with the monetary gains from the game. The dictator thus faces a trade-off between securing a high monetary payoff and being perceived as fair. Being observed by an audience reinforces those image concerns and thereby decreases the dictator's expected overall payoff, leading to a weaker preference for that role.

As hypothesized, we find that dictator offers are higher in the *public* treatment. The main driving factors are one's own fairness concerns and awareness of one's image concerns. Correspondingly, the willingness to pay to become the dictator is significantly lower in the *public* treatment compared to the *private* treatment. Here again, one's own fairness and awareness of one's own image concern explain the results. Additionally, in the *private* treatment, high expectations about others' fairness decrease auction bids. High expectations increase the value of the recipient role, thereby decreasing bids. With the public announcement, this effect vanishes as the expected generosity shifts upward overall.

The paper proceeds as follows. Section 2 discusses the most closely related literature. Section 3 presents the experimental design, a simple model, and our hypotheses. The data analysis and results are shown along with two robustness checks in Section 4. Section 5 discusses the results and concludes.

3.2 Related Literature

We build on an extensive literature on social observability in decision-making, specifically in the dictator game. Many experimental studies have found that, *ceteris paribus*, decreasing dictators' anonymity increases their generosity (e.g. Hoffman et al. (1994, 1996); Bohnet and Frey (1999b); Charness and Gneezy (2008); Franzen and Pointner (2012); Alevy et al. (2014)). The consensus of the papers mentioned above is that there is a general positive relationship between observability and generosity. Engel (2011)'s meta-study finds that if dictators are identified, they are less likely to give nothing and more likely to give more than half of the pie. However, some studies find null (Bolton et al., 1998; Johannesson and Persson, 2000; Frohlich et al., 2001; Barmettler et al., 2012; Dreber et al., 2013) or even contradictory results (Dufwenberg and Muren, 2006; Rankin, 2006).

Frey and Bohnet (1995); Bohnet and Frey (1999a) systematically vary the social distance between dictator and recipient similarly as our experiment, but it significantly differs concerning the timing of information provision. In their experiments, subjects stand up and look at other participants *before* playing dictator games, such that dictators gather information about their opponent before making a choice. Image concerns and fairness preferences are likely to vary with personal characteristics such as gender, ethnicity, age, and so on. With the ex post publication of the decision, we can retain recipients' anonymity while maximizing the dictators' visibility. We believe our ex post revelation procedure is more conservative in separating image concerns from other confounding factors.

An even more closely related experimental design is presented by Dufwenberg and Muren (2006), who hand out payments for a dictator game either in front of an entire classroom or in private. The authors find that significantly *less* is shared when payments are made in public. However, Dufwenberg and Muren (2006) cannot exclude that the effect is driven by the specific subject pool consisting of undergraduate economics students, who arguably follow different norms and expectations of how a dictator ought to behave, especially in the presence of the instructors teaching them game theory. As our sample consists of fewer than one-third of economics and business students, and the experiment is conducted outside the classroom and lecture context, we do not expect these experimenter-demand factors to have a significant influence in our case.

Jones and Linardi (2014) show that the "wallflower effect" draws (some) subjects towards behaving like the average group member due to an aversion against standing out. Therefore, visibility might increase pro-social behavior only when the group's level of pro-sociality is expected to be high. This aligns with the findings of Andersson et al. (2020); Bolton et al. (2021).

For our theoretical analysis, we rely on the work of Andreoni and Bernheim (2009), who propose a simple inequity aversion model and extend it to incorporate social image concerns. Thereby, a player's behavior is not only driven by her own preference to act socially but also by the impression her action leaves on others. In such a signaling game, players increase their dictator offers when their actions are visible. Andreoni and Bernheim (2009) explain pooling at the equal split of the surplus and show in an experiment how image concerns drive such pooling behavior. Our argument concerning role preferences in the dictator game builds on the signaling rationale introduced by Andreoni and Bernheim (2009). In contrast to their model, we treat image concerns as additional (non-monetary) costs rather than benefits. This assumption is in line with findings by Dana et al. (2006), who show that subjects have a positive willingness to pay to avoid being in the situation of a dictator. Broberg et al. (2007) estimate exit reservation prices in a modified replication of Dana et al. (2006) and find that approximately two-thirds of the subjects are happy to quietly leave a dictator game with a smaller amount than the surplus to be divided. Andreoni et al. (2017) present a field experiment with similar findings. Our contribution in this line of research is to quantify the influence of social control on the willingness to pay for the dictator role.

We quantify preferences over the two roles in the dictator game via a second-price auction. To the best of our knowledge, no experiment has used an auction in such a way to allocate the roles in a dictator game before. There are two experiments using auctions to sell *participation with pre-defined roles* in the ultimatum game: Güth and Tietz (1986) via a second-price auction and Shachat and Swarthout (2013) via an English clock auction. Subjects are randomly allocated to either the potential proposer or the potential responder pool and can then bid on entering the game. Güth and Tietz (1986) find that proposers bid on average almost twice as much as responders. Shachat and Swarthout (2013) find that auction prices often reflect beliefs inconsistent with Nash equilibria (notably in games with only monetary payoffs). Our experiment differs in that both papers mentioned above use an ultimatum game and cannot directly measure the intensity of the preference for being the first mover, since roles are predetermined.[2] Furthermore,

[2]The willingness to pay for one role over the other can only be implicitly estimated as the difference in the willingness to pay between the potential proposers and responders and by assuming that the value of the game itself is constant across the two roles. The difference measure is not accurate when there are

the auction may create an entitlement effect, as not every subject can participate in the subsequent ultimatum game. Entitlement has been found to significantly influence behavior in several studies (Hoffman et al., 1994; Frohlich et al., 2004; García-Gallego et al., 2008; Kassas and Palma, 2019). We therefore interpose a randomization between the auction and the bargaining situation to minimize such entitlement effects: a proposer cannot know for sure how she obtained her first-mover position and thus should not feel a stronger entitlement to the surplus than the responder.

To conclude, the monetary attractiveness of the dictator role seems indisputable from a game-theoretic perspective, especially for selfish players. However, behavioral insights from other studies on social image and inequity aversion contest this presumption, which has not been rigorously tested to date. Our primary treatment variable, social image, has yielded ambiguous results in previous studies, suggesting that additional systematic evidence is needed to clarify the exact mechanism by which social context influences bargaining behavior.

3.3 Experimental Design

3.3.1 The Games Played

The experiment consists of three parts as depicted in Figure 3.1. In Part 1, subjects play a dictator game (DG 1) using the strategy method, unaware whether they will be the dictator or the recipient. Every player makes the hypothetical choice of how to distribute 40 points between herself and another unknown player. Which players are paired and whose allocation is implemented is randomly chosen. Subjects learn about the realized payoff from this part at the very end of the entire experiment, and the payoffs from DG 1 are never made public. The strategy method allows us to use the number of points received in the first dictator game as a measure of each subject's fairness preferences in the absence of image concerns arising from public payment. During Part 1, subjects are aware that other parts will follow, but remain ignorant of the exact games to be played later. All actions are taken in a one-shot manner. Additionally, we elicit subjects' first-order beliefs in DG 1 in an incentivized manner, similar to Krupka and Weber (2013).

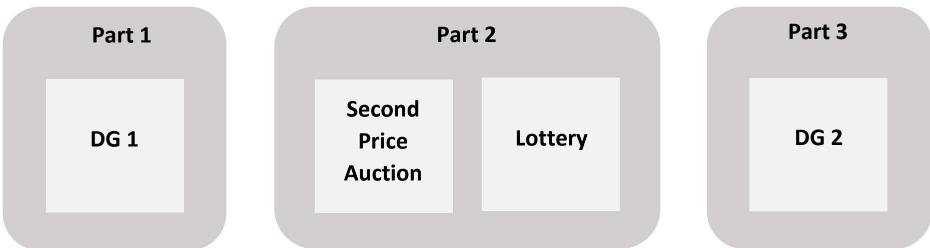


Figure 3.1: Experimental Set-Up

In Part 2, subject pairs are matched for the rest of the experiment. Subjects learn that another dictator game (DG 2) will follow. Both players $i = 1, 2$ in a pair privately place a bid in a sealed-bid second-price auction to increase the probability of becoming the dictator by doing

interactions between the game value and the role preference.

the following: they each receive an endowment of 100 points and submit a bid $b_i \in [0, 100]$. The higher bidder in a pair gets a lottery ticket that gives this player the dictator role in DG 2 with 90% probability and the recipient role with 10%. Vice versa, the player with the lower bid wins the dictator role with a probability of 10% and the recipient role with a probability of 90%. The auction winner pays the loser's bid to the experimenter, and the loser keeps her entire endowment from Part 2. We use this strategy-proof mechanism to gauge subjects' willingness to pay in an incentive-compatible way, as bidding one's true valuation is a weakly dominant strategy. In case of a draw, the auction winner is randomly determined by the computer with equal chances and pays her bid.

The lottery draw is carried out secretly by the computer and independently for each pair. At the end of Part 2, subjects are informed about the role assignment after the lottery, but neither learn who won the auction, nor the price the winner pays for the ticket. The lottery is meant to limit spillovers of confounding factors between the auction and DG 2. First, learning about the opponent's bid and thereby their willingness to pay for the dictator role could be informative about their preference type. Second, winning the auction could trigger the aforementioned entitlement effects. Via the lottery, we keep players unaware of whether their bid got them the role they are in, or whether it was sheer luck. However, a 90% winning probability keeps the noise to a minimum. It is still high enough to make bidding attractive.[3]

After the role assignment, the dictator game DG 2 is played in Part 3, this time with direct response instead of the strategy method.[4] The dictator splits an additional surplus of 100 points between themself and the recipient. The procedures of the experiment, as described up to this point, are the same in both treatments. Our treatment variation is used to implement the dictator's allocation, as described in the next section.

3.3.2 The Treatments

We have two treatment conditions, a *public* and a *private* one. In the *public* treatment, at the very end of the session, dictators are required to step out of their cubicle, such that all other participants can see them. The experimenter reads aloud the list of cubicle numbers along with the players' sharing decision made in DG 2. The recipients see their corresponding dictator's cubicle number and share the decision on their screen, and can thus identify the person they were paired with, as well as learn about all other dictators' decisions. The recipients do not stand up and are not identified. The instructions for the *public* treatment describe this procedure in detail to make the treatment salient to subjects when they make their decisions in Parts 2 and 3.

The *private* treatment does not include the announcement. Instead, the experiment ends directly after Part 3, and subjects learn their payouts only from their computer screens. The crucial difference, therefore, is the visibility of the dictators' decision: the *public* treatment effectively heightens participants' image concerns. The payment is carried out privately in both treatments. Individual payoffs consist of the sum of points from the dictator game DG 1 in Part 1, the

[3]To the best of our knowledge, such a randomization device has not been used in dictator games, but other studies have implemented comparable mechanisms to circumvent issues of self-selection, e.g. Dal Bó et al. (2010) use a lottery with 90% and 10% probabilities to assign players to different games.

[4]Loewenstein (2000) explains how visceral factors such as fear are difficult to anticipate, while their actual appearance tremendously affects an individual's decision-making. We chose the direct response method because our main motive of interest in DG 2 is a dictator's image concerns, which essentially coincide with a fear of stigmatization. By direct response, we maximize the likelihood of suspending our subjects in this emotion. Brosig et al. (2003) find a significant change in behavior for subjects under the direct response method when such non-monetary motives play a role.

remainder of the auction endowment in Part 2, the dictator game DG 2 in Part 3, and a potential reward for one randomly chosen belief question.

3.3.3 Theoretical Considerations and Hypotheses

The model is a simplistic reproduction of the experimental design. We attempt to abstract the complex experimental situation to make it mathematically tractable within a behavioral game-theoretic model. Our model is primarily inspired by the seminal work of Andreoni and Bernheim (2009) and Bénabou and Tirole (2006) on image concerns.
We can illustrate the main mechanism behind social-image concerns and first-mover preferences in negotiations by applying a second-price auction as an allocation device for an ensuing dictator game. We then analyze strategic bidding behavior across treatments and conclude with testable hypotheses about the experiment's outcome.

The situation is modeled as a sequential game with incomplete information. Two players are matched and compete for the role of the dictator. Both players i, with $i = 1, 2$, receive an initial endowment, and decide which part to invest in a dichotomous choice second-price auction. The auction sells the dictator role to the highest bidder. The winner pays the second-highest bid and immediately becomes dictator, and the loser becomes the recipient and pays nothing [5]. In the event of a draw, a 50:50 lottery determines the winner. Once the roles are allocated, the player who becomes the dictator decides between a high and a low offer, with or without an observing audience. The audience itself makes no decision, but forms a belief about a player's type. Because some players are concerned about their social image, the audience's beliefs enter their utility functions, and the game becomes a signaling game. The signal is the players' dictator offer. Depending on the audience's ability to observe the player's offer, controlled by the treatment condition, image concerns affect behavior, and different equilibria emerge.

In the following, we recount the sequence of events.

1. Nature matches two players and chooses their types. There are two player types $t^i_\lambda = (\theta, \mu)$, with $\lambda = 1, 2$, that differ on the two type dimensions fairness preference (θ) and image concern (μ). The first type $t^i_1 = (1, 0)$ is only fairness concerned ($\theta_i = 1$) and not image-concerned ($\mu_i = 0$). The common prior on this type is $Pr(t^i = t^i_1) = p$. The second type $t^i_2 = (0, \frac{2}{3})$ is partially monetary ($1 - \theta_i - \mu_i = \frac{1}{3}$) and image concerned ($\mu_i = \frac{2}{3}$). The common prior on this type is $Pr(t^i = t^i_2) = 1 - p$. Each player i knows their own type, but not their opponent j's type, with $j = 1, 2$ and $j \neq i$. The audience does not know the players' types but shares the same common prior belief about the distribution of types.
2. The players simultaneously choose between a high b_h and low b_l bid, with $b_h > b_l$.
3. The player with the higher bid wins the auction, pays the amount of the lower bid, and immediately becomes dictator. In the event of a draw, nature draws a winner by lot, who becomes dictator and pays their own bid.
4. As the dictator, a player chooses between a high o_h and a low o_l offer, with $o_h > o_l$, in the dictator game. If a player becomes the recipient, the game ends for them, and they receive what their opponent j offered.

[5] Compared to the experiment, we do not use a lottery and instead directly assign the dictator role to the auction winner. The utility functions in our model neither include a motive for the entitlement effect, nor a reason to behave differently in the dictator game depending on the expected preference type of the opponent.

5. The audience observes the dictator's offer, forms posterior beliefs $Pr(t^i = t^i_\lambda | o_k)$ and thereby calculates a conditional expectation on the player's fairness type $E[\theta | o^i_k]$, with $k = h, l$. Exemplary for player 1 being the dictator, if the audience observes a high offer (o_h) from player 1, they are in information set $I^A_{o_h}$ of the following game tree. If they observe a low offer (o_l), they are in information set $I^A_{o_l}$. The audience cannot distinguish between the player's type and their bid.

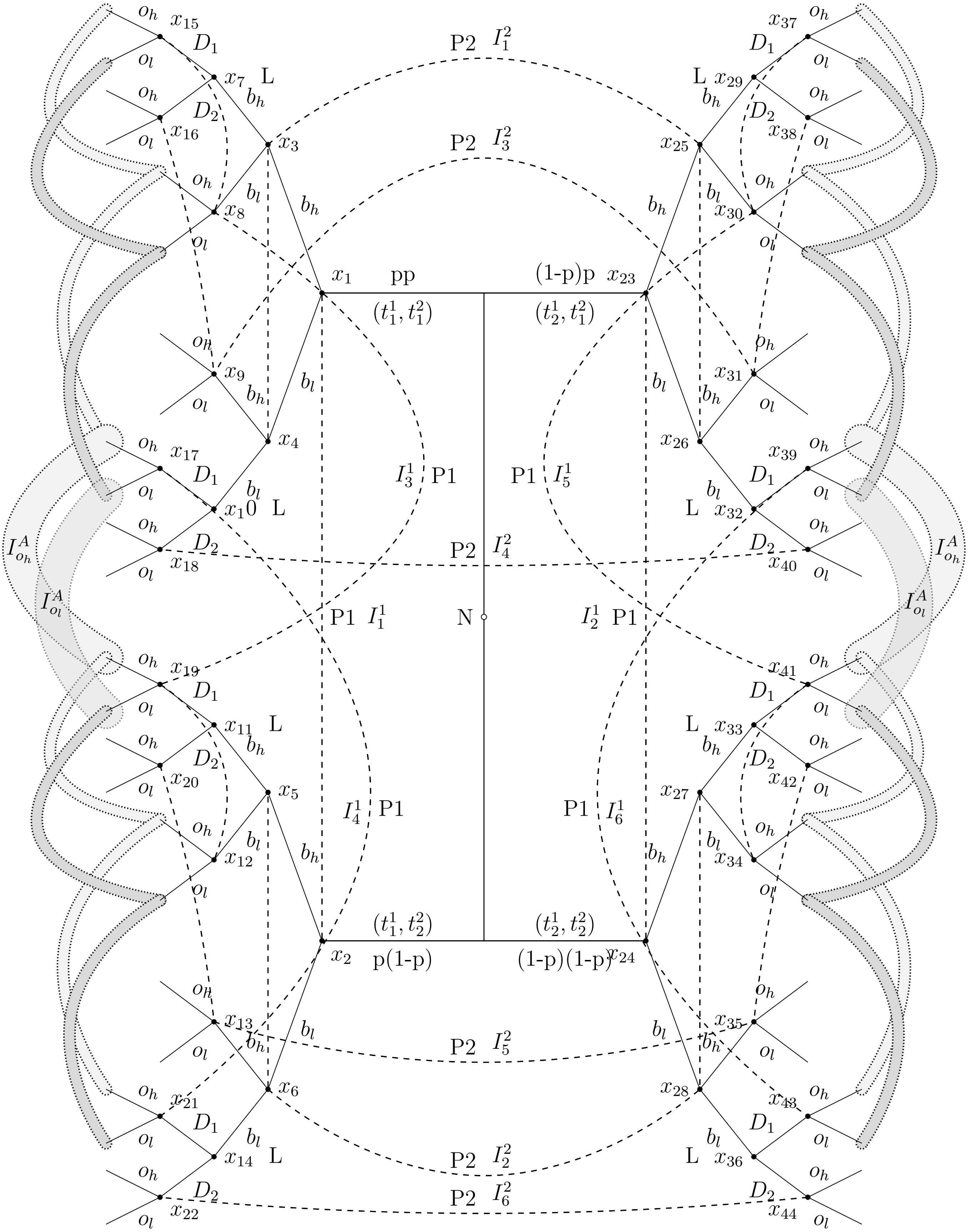


Figure 3.2: Game Tree Representation

The illustration in figure 3.2 presents the entire game tree with the two players' actions and information sets, as well as the audience's information sets $I^A_{o_h}$ and $I^A_{o_l}$ observing player 1 acting as the dictator[6]. P1 marks player 1 and P2 marks player 2. The 44 decision nodes are divided into six information sets for each player, all of which include more than one decision node.

A strategy s_i of a player i prescribes an action in each information set of the player. Each player i has six information sets I^i_τ, with $\tau = 1, 2, ..., 6$. Player i can either be of a t^i_1 or t^i_2 type. Depending on their own type they are either in I^i_1 or I^i_2. In each set, they cannot differentiate between the other player j's type. Both information sets include decision nodes for the second-price auction, in which the player chooses between a high (b_h) and a low (b_l) bid.

We formalize the decision by the concept of behavioral strategies, a probability distribution over b_h and b_l, and label the probability that player i decides for the high bid in information set I^i_τ, with $\tau = 1, 2$, as $\beta_i(I^i_\tau)$, with $\beta_i \in [0, 1]$. Player 1 choosing the high bid in I^1_1 expresses as $\beta_1(I^1_1) = 1$, while choosing the low bid in I^1_2 as $\beta_1(I^1_2) = 0$. Any β_i between 0 and 1 corresponds to a mixed strategy, which we neglect in the analysis.

If player 1 is a t_1-type, they are in I^1_1. Next, they find themselves either in I^1_3 or I^1_4. In I^1_3 they have become the dictator and chose a high bid. They still do not know which type they are facing, nor which decision of player 2 lead to this information set. Therefore, I^1_3 contains four decision nodes and can therefore be the consequence of four different histories. Two in which their opponent is a t_1-type and two in which their opponent is a t_2-type. For each of these two, in one node the opponent chose a high bid, and they became the dictator by chance, and in the other, the opponent chose a low bid, and they won the auction. In I^1_4 they chose a low bid and became dictator. They thereby know that their opponent also chose a low bid and that they got the role by chance. In both information sets I^1_3 and I^1_4, they decide between a high offer o_h and a low offer o_l.

Again, we formalize the decision by the concept of behavioral strategies, a probability distribution over o_h and o_l. We label the probability that player i chooses the high offer in in information set I^i_τ, with $\tau = 3, 4$, as $\alpha_i(I^i_\tau)$, with $\alpha_i \in [0, 1]$. Player 1 choosing the high offer in I^1_3 expresses as $\alpha_1(I^i_3) = 1$, while choosing the low offer in I^i_4 as $\alpha_i(I^i_4) = 0$. Any α_i between 0 and 1 corresponds to a mixed strategy, which we neglect in the analysis.

If player 1 is a t_2-type they are in I^1_2. After their bidding decision, they either find themselves in I^1_5 or I^1_6. A similar logic as before applies to these two information sets, and consequently we label the respective decisions $\alpha_1(I^1_5)$ and $\alpha_1(I^1_6)$.

For player 2, the information sets follow the same logic. The only difference stems from the visualization chosen for the game tree. Although the auction decisions are simultaneous, we illustrated them in an order due to limitations in the game tree logic. The impact of visually ordering is eliminated by extending the two information sets I^2_1 and I^2_2 by two decision nodes each, thereby directly incorporating the uncertainty about the opponent's type into the opponent's auction choice.

[6]We omitted the audience's information sets for player 2 acting as dictator for reasons of overview.

We can now summarize that for a player i a strategy contains a prescribed action in each information set and therefore we write $s_i = (\beta_i(I_1^i), \beta_i(I_2^i), \alpha_i(I_3^i), \alpha_i(I_4^i), \alpha_i(I_5^i), \alpha_i(I_6^i))$. The first two entries concern the information sets I_1^i and I_2^i and prescribe a behavioral strategy for the high (b_h) or low (b_l) bid in the auction, depending on the type. The last four entries are concerned with the information sets I_3^i, I_4^i, I_5^i and I_6^i and prescribe a behavioral strategy over the high (o_h) or low (o_l) offer in the dictator game, following a certain combination of types and auction behavior.

To develop a player's utility function, we start by collecting its components from the dictator game. Player i as the dictator has the following utility function in the information sets I_τ^i with $\tau = 3, 4, 5, 6$.

$$\begin{aligned} U_i^D(\alpha_i(I_\tau^i); \sigma, t_\lambda^i(\theta_i, \mu_i)) = \alpha_i(I_\tau^i)\Big[&(1 - \theta_i - \mu_i)(S - o_h) - \theta_i(o^{eq} - o_h) - \mu_i\sigma(\bar{\theta} - E[\theta|o_h])\Big] \\ +(1 - \alpha_i(I_\tau^i))\Big[&(1 - \theta_i - \mu_i)(S - o_l) - \theta_i(o^{eq} - o_l) - \mu_i\sigma(\bar{\theta} - E[\theta|o_l])\Big] \end{aligned} \tag{3.1}$$

The dictator's utility potentially consists of three motives. The financial motive $(1-\theta_i-\mu_i)(S-o_k)$, the fairness motive $-\theta_i(o^{eq} - o_k)$ and the image concern $-\mu_i\sigma(\bar{\theta} - E[\theta|o_k])$, with $k = h, l$. The financial motive is weighted as the residual of fairness and image concerns and consists of the surplus S minus the dictator offer o_k. The fairness motive is weighted by the player's fairness preference θ_i and consists of the difference between the equal outcome $o^{eq} = \frac{1}{2}S$ and the dictator offer o_k. Any difference decreases a fairness-concerned player's utility. For a high offer (o_h), the difference is minimized. The image concern is weighted by the player's image preference, μ_i, and the signal observability parameter, σ. Between the treatments, σ changes. In the *public* treatment, the audience can perfectly observe the dictator's offer and $\sigma = 1$. In the *private* treatment, the audience cannot observe the dictator's offer and therefore σ must be lower. We set $\sigma = \frac{1}{10}$, because we do not want to exclude the possibility that a player still experiences a reduced, but still existent image concern. The concern itself consists of the difference between the highest fairness preference possible, namely $\bar{\theta} = 1$ and an external audience's conditional expectation after observing a given dictator offer $E[\theta|o_k]$. Player i knows that the audience cannot distinguish between their type and their action in the auction. It only observes a high or low offer of player i. The player's image, therefore, depends on their action and the audience's posterior belief formed after observing it. An action implying a high fairness type is favorable. In this modeling approach, we assume image concerns to enter the player's objective function as costs. We argue that failing to act in accordance with the fairness norm and implementing an unequal allocation casts a bad light on an individual. An alternative way would be to model image concerns as benefits in the objective function. Under this approach, an individual can

shine in light of behavior proscribed by the norm. In a robustness check, we find arguments for our cost approach.
The utility functions for the two types reduce to the following expressions.

$$U_i^D(\alpha_i(I_\tau^i);\sigma,t_1^i) = \alpha_i(I_\tau^i)\Big[-(o^{eq}-o_h)\Big] + (1-\alpha_i(I_\tau^i))\Big[-(o^{eq}-o_l)\Big] \tag{3.2}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\tau^i);\sigma,t_2^i) =& \alpha_i(I_\tau^i)\Big[\frac{1}{3}(S-o_h) - \frac{2}{3}\sigma(\bar{\theta}-E[\theta|o_h]\Big] \\ &+(1-\alpha_i(I_\tau^i))\Big[\frac{1}{3}(S-o_l) - \frac{2}{3}\sigma(\bar{\theta}-E[\theta|o_l]\Big] \end{aligned} \tag{3.3}$$

We see that the t_1-type is only interested in equal outcomes, while the t_2-type weighs financial motives against image concerns in a 1:2 ratio and is completely disinterested in equal outcomes.

As the recipient, player i has the following utility function, depending on player j's dictator decision ($j \neq i$), in the information sets I_τ^j for $\tau = 3, 4, 5, 6$.

$$\begin{aligned} U_i^R(\alpha_j(I_\tau^j),\sigma,t_\lambda^i(\theta_i,\mu_i)) = & \alpha_j(I_\tau^j)\Big[(1-\theta_i-\mu_i)(S-o_h)-\theta_i(o^{eq}-o_h)\Big] \\ &+(1-\alpha_j(I_\tau^j))\Big[(1-\theta_i-\mu_i)(S-o_l)-\theta_i(o^{eq}-o_l)\Big] \end{aligned} \tag{3.4}$$

The recipient cannot actively influence their utility. The image concern is not present, because they are not active and an audience cannot directly learn anything about their fairness preference. Plugging in the respective types' preferences gives us the following expressions.

$$U_i^R(\alpha_j(I_\tau^j),\sigma,t_1^i) = \alpha_j(I_\tau^j)\Big[-\theta_i(o^{eq}-o_h)\Big] + (1-\alpha_j(I_\tau^j))\Big[-\theta_i(o^{eq}-o_l)\Big] \tag{3.5}$$

$$U_i^R(\alpha_j(I_\tau^j),\sigma,t_2^i) = \alpha_j(I_\tau^j)\Big[\frac{1}{3}(S-o_h)\Big] + (1-\alpha_j(I_\tau^j))\Big[\frac{1}{3}(S-o_l)\Big] \tag{3.6}$$

Again, we see how the t_1-type is only interested in equal outcomes and the t_2-type only considers their financial gains as image concerns play no role for the recipient.

We can use the dictator game utilities to describe the utility functions at the second-price auction stage. In the second-price auction, player i competes with player j. Both choose between a high b_h and a low b_l bid. The objective function for a t_1-type player i is the following.

$$
\begin{aligned}
U_i^A(s_i, s_j; \sigma, t_1^i) = 1 + p\Big[\\
&\beta_i(I_1^i)\Big(\beta_j(I_1^j)\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big) \\
&\quad +(1-\beta_j(I_1^j))\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big) \\
&+(1-\beta_i(I_1^i))\Big(\beta_j(I_1^j)\Big(U_i^R(\alpha_i(I_3^j), \sigma, t_1)\Big) \\
&\quad +(1-\beta_j(I_1^j))\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big)\Big)\Big] \\
+(1-p)\Big[\\
&\beta_i(I_1^i)\Big(\beta_j(I_2^j)\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big) \\
&\quad +(1-\beta_j(I_2^j))\Big(U_i^D(\alpha_i(I_3^j); \sigma, t_1) - b_l\Big)\Big) \\
&+(1-\beta_i(I_1^i))\Big(\beta_j(I_2^j)\Big(U_i^R(\alpha_i(I_5^j), \sigma, t_1)\Big) \\
&\quad +(1-\beta_j(I_2^j))\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_i(I_6^i), \sigma, t_1)\Big)\Big)\Big]
\end{aligned}
\tag{3.7}
$$

The objective function for a t_2-type player i differs with respect to the information sets, in which player i makes their decisions.

$$
\begin{aligned}
U_i^A(s_i, s_j; \sigma, t_2^i) = 1 + p\Big[\\
&\beta_i(I_2^i)\Big(\beta_j(I_1^j)\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big) \\
&\quad +(1-\beta_j(I_1^j))\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_1) - b_l\Big)\Big) \\
&+(1-\beta_i(I_2^i))\Big(\beta_j(I_1^j)\Big(U_i^R(\alpha_i(I_3^j), \sigma, t_1)\Big) \\
&\quad +(1-\beta_j(I_1^j))\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big)\Big)\Big] \\
+(1-p)\Big[\\
&\beta_i(I_2^i)\Big(\beta_j(I_2^j)\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big) \\
&\quad +(1-\beta_j(I_2^j))\Big(U_i^D(\alpha_i(I_5^j); \sigma, t_1) - b_l\Big)\Big) \\
&+(1-\beta_i(I_2^i))\Big(\beta_j(I_2^j)\Big(U_i^R(\alpha_i(I_5^j), \sigma, t_1)\Big) \\
&\quad +(1-\beta_j(I_2^j))\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_i(I_6^i), \sigma, t_1)\Big)\Big)\Big]
\end{aligned}
\tag{3.8}
$$

In the second-price auction, both players know their own type but bid under uncertainty about their opponent j's type. For each type, they choose a behavioral strategy $\beta_i(I_\lambda^i)$, with $\lambda = 1, 2$, that maximizes their expected payoff after receiving a standardized endowment of one. The expected payoff depends on the distribution of the opponent's type $(p, 1-p)$, their chosen bidding behavior $\beta_j(\cdot)$, player i's own bidding behavior $\beta_i(\cdot)$ and both players' behavior in the ensuing dictator game $\alpha_i(\cdot)$ and $\alpha_j(\cdot)$. The different combinations of behavior result in player i either loosing, winning or arriving at a draw in the auction.

The first set of square brackets in either type's objective function is multiplied by the probability of facing a t_1-type opponent j. Inside the brackets, we find the payoff of facing such a type, depending on one's own and the opponent's behavior. The second set of square brackets is multiplied by the probability of facing a t_2-type opponent j. Inside the brackets, we find the payoff of facing such a type.

The player with the higher bid wins and pays the lower bid. Whenever there is a draw, a 50:50 lottery determines the winner. The payoffs of becoming the dictator U_D or recipient U_R depend on the behavior the respective player i or j chooses in the ensuing dictator game.

The audience's posterior beliefs affect the behavior of an image-concerned player i through conditional expectations over their fairness preference. An image concerned player takes on the perspective of an outside audience and integrates their judgment into their own utility function. A high fairness preference is valued. An audience has a prior belief on the distribution of types $(p, 1-p)$, with $Pr(t=t_1)=p$ and $Pr(t=t_2)=1-p$. Based on this prior and the observed dictator i's behavior o^i_k, they form a posterior belief $Pr(t=t^i_\lambda|o^i_k)$, with $k=h,l$ labeling a high or low offer or bid, $\lambda=1,2$ labeling the type and $i=1,2$ labeling the player. Weighting each type's fairness preference by its corresponding posterior yields the conditional expectations.

$$\begin{aligned} E[\theta|o^i_h] &= 1\cdot Pr(t^i_1|o^i_h)+0\cdot Pr(t^i_2|o^i_h)=Pr(t^i_1|o^i_h) \\ E[\theta|o^i_l] &= 1\cdot Pr(t^i_1|o_l)+0\cdot Pr(t^i_2|o^i_l)=Pr(t^i_1|o^i_l) \end{aligned} \tag{3.9}$$

To calculate the conditional expectations, it suffices to compute the posteriors $Pr(t^i_1|o^i_h)$ and $Pr(t^i_1|o^i_l)$, as the fairness preference of a t_2-type is zero, and thereby the respective posteriors cancel out.

$$Pr(t^i_1|o^i_h)=\frac{Pr(o^i_h,t^i_1)}{Pr(o^i_h)}=\frac{Pr(o^i_h,t^i_1)}{Pr(o^i_h,t^i_1)+Pr(o^i_h,t^i_2)} \tag{3.10}$$

$$Pr(t^i_1|o^i_l)=\frac{Pr(o^i_l,t^i_1)}{Pr(o^i_l)}=\frac{Pr(o^i_l,t^i_1)}{Pr(o^i_l,t^i_1)+Pr(o^i_l,t^i_2)} \tag{3.11}$$

In the following analysis, we solve the game for its Perfect Bayesian equilibrium. This solution concept for dynamic games of incomplete information requires that players' posterior beliefs on the equilibrium path are formed in accordance with Bayes' rule and are consistent with equilibrium strategies. In appendix A the calculation of posteriors is summarized.

The game is a dynamic game with incomplete information. We therefore determine its Perfect Bayesian equilibria. It is a symmetric game in the way that both players have the same set of strategies and objective functions. Therefore, we determine symmetric equilibria $s = (s_i, s_j)$, in which both players play the same strategy $s_i = s_j$. Therefore, it is sufficient to show that a strategy for player i is part of an equilibrium, as the same argument is true for player j. Further, we will concentrate on pure strategies. To simplify the analysis, we set the following parameter values. We standardize the surplus to one $S = 1$. This leads to the equal outcome being $o^{eq} = \frac{1}{2}$. Further, we set the high offer to the equal outcome $o_h = o^{eq} = \frac{1}{2}$, thereby diminishing any inequality. Finally, the low bid is set at zero, $b_l = 0$.
A first inspection of the players' incentives results in the following propositions about possible equilibria.

Proposition 1. *Independent of the treatment, the t_1-type always prefers o_h over o_l in both information sets I^i_κ, with $\kappa = 3, 4$.*

From proposition 1, we learn that any potential equilibrium strategy s_i includes the t_1-type choosing the high offer o_h. This reduces the set of equilibrium candidate strategy profiles. Also, because the t_1-type never chooses the low offer o_l the posterior belief $Pr(t = t^i_1|o^i_l)$ is zero whenever o_l is on the equilibrium path.

Proposition 2. *Depending on the relation between image premium $E[\theta|o_h] - E[\theta|o_l]$ and financial incentives $o_h - o_l$ the t_2-type either chooses the high o_h or low o_l offer in both information sets I^i_κ, with $\kappa = 5, 6$.*

From propositions 1 and 2, we conclude that there are two different types of equilibrium candidates. In a *fully separating equilibrium*, the t_1-type chooses o_h in both information sets and the t_2-type chooses o_l. The types fully reveal themselves by their actions. In these equilibria, behavior is a perfect signal, and the posterior expresses certainty. In a *fully pooling equilibrium*, the t_1-type again chooses o_h in both information sets, but the t_2-type chooses o_h as well. In this case, the posterior must be carefully calculated in coherence with Bayes' rule, and the action is an imperfect signal on the dictator's type.

Proposition 3. *In the public treatment, there exists no separating equilibrium. In the private treatment there only exist separating equilibria, if the financial incentives are high enough ($o_l < \frac{3}{10}$)*

For a t_2-type, the monetary incentives must be strong enough to choose the low offer in the *private* treatment. In the *public* treatment, there is no feasible financial incentive that makes the t_2-type prefer the low offer. In order to provide conditions for both types of equilibria to emerge, we apply the additional assumption that $o_l < \frac{3}{10}$ and thereby assume significant financial incentives given our earlier parameter assumptions.

Result 1. *In the private treatment, there potentially exist three different kinds of separating equilibria, depending on the combination of financial incentives o_l, the cost of a high bid b_h, and the type distribution $(p, 1-p)$.*

1) $s = (s_1, s_2)$ *with* $s_i = (1,1,1,1,0,0)$, $i = 1,2$

- *For* $b_h < \frac{1}{20}(3 - 10o_l)$ *and* $p < \frac{1-2b_h-2o_l}{1-2o_l}$
- *For* $\frac{1}{20}(3 - 10o_l) < b_h < \frac{4}{15} - \frac{2}{3}o_l$ *and* $p < \frac{8-30b_h-20o_l}{5-10o_l}$

2) $s = (s_1, s_2)$ *with* $s_i = (1,0,1,1,0,0)$, $i = 1,2$

- *For* $b_h > \frac{-11+52o_l-60o_l^2}{-14+20o_l}$ *and* $\frac{8-20o_l}{5+30b_h-10o_l} \leq p \leq \frac{1-2o_l}{2+2b_h-4o_l}$

3) $s = (s_1, s_2)$ *with* $s_i = (0,1,1,1,0,0)$, $i = 1,2$

- *For* $b_h > \frac{1}{2} - o_l$ *and* $\frac{8-30b_h-20o_l}{5-30b_h-10o_l} \leq p < 1$

We find three potential separating equilibria in the *private* treatment. All three differ with respect to the types' bidding behavior in the second-price auction. In 1), both types choose a high bid. They do so for relatively small values of the parameters b_h and o_l over a broad range of probabilities. For high financial incentives - low o_l and at a low price - low b_h, both types try to increase their chance of becoming the dictator. In 2), only the t_1-type bids high. This equilibrium exists only for a rather narrow range of low probabilities and high costs b_h. If there are few t_1-types, the high likelihood of inequality pushes the few t_1-types to become dictators, albeit at a high price, a high b_h. In 3) only the t_2-type chooses the high bid. This equilibrium exists for an intermediate range of b_h and probabilities. Chances are high that the high bid will be effective as the prevalent t_1-types choose the low bid.
From Engel (2011), a meta-study on the dictator game, we know that about 17% of dictators choose the equal split and thereby completely fall into the t_1-type category. About another 25% give between half and about one-third of the surplus to the recipient. From Dana et al. (2007), we know that about 50% of dictators choose ignorance to achieve a favorable outcome for themselves. Our model's dichotomous type structure simplifies reality by splitting the entire population into two types. We therefore assume that a proportion of at least 40% fall into the t_1-type category and thereby $p \geq \frac{4}{10}$. This assumption rules out equilibria of the kind in 2). We are left with separating equilibria 1) and 3).

Result 2. *In the public treatment, there may be a pooling equilibrium.*

4) $s = (s_1, s_2)$ *with* $s_i = (0,0,1,1,1,1)$, $i = 1,2$

- *For* $p > \frac{1}{4}(1 - 2o_l)$

We find one potential pooling equilibrium in the *public* treatment. In 4), both types choose the low bid for any feasible probability p. As both choose the high offer, they expect the same payoff whether they become dictator or not. Therefore, they have no reason to invest in this role. Also, the audience cannot distinguish between the two types by their action. Any deviation would lead to an immediate image cost for the t_2-type.

These two results highlight a specific dynamic between our *public* and *private* treatments. Observing this dynamic in the the data can be interpreted as evidence for our theory. We summarize that, across all circumstances, image concerns lead to lower average offers and higher average bids in the *private* treatment than in the *public* treatment. To underscore the explanatory significance of our theory, in appendix A we conduct two robustness checks. We can rule out the possibility that the same phenomenon stems from the t_1-type facing either of the two alternative preference types. Instead, the observed dynamic is solely dependent on image-related concerns that arise as costs.

In a first alternative scenario, we test what happens between treatments when the t_1-type is a purely selfish, financially motivated player. We find that in both treatments, there exist only separating equilibria of the kind in 1) and 3). The *public* treatment does not have the same competition-diminishing effect on the auction as under image concerns as costs.

In the second alternative scenario, we test what happens between treatments if the t_1-type takes on an image-concerned player, for whom image concerns are a benefit. This type intends to shine by signalling a high fairness concern. In this case, instead of the pooling equilibrium where both types bid low, a new pooling equilibrium arises. The image concerned type places a high bid to increase their chance of becoming the dictator. Image concerns occuring as benefits always incentivize the image-concerned type to bid high and jump at the chance to shine.

The earlier-described dynamic is thus specific to image concerns that arise as costs and can therefore count as evidence for our theory. We thereby arrive at two simple, testable hypotheses for our experiment.

Hypothesis 1. *Average dictator offers are higher in the public treatment.*

The only equilibrium in the *public* treatment is a pooling equilibrium. Both types choose the high offer: one in line with their fairness preference and the other to signal their virtue. Although in the experiment there is a whole continuum of types, we expect similar dynamics.

Hypothesis 2. *The average auction bid is higher in the private treatment.*

In the *private* treatment, only the two separating equilibria of the kind as in 1) and 3) are possible. In 1) both types choose the high bid, and in 3) the t_2-type chooses the high bid. In either case, compared to the *public* treatment, more players choose the high bid, thereby increasing the average bid.

3.4 Analysis

This section presents the analysis of the experimental data in accordance with the hypotheses derived from the theoretical discussion in Section 3. To this end, we first introduce the key variables needed to test the hypotheses. Second, the protocol of the experimental sessions is presented along with the characteristics of the subject pool. Afterwards, descriptive statistics and results of regression analyses are shown to test the hypotheses and to assess potential transmission channels between the *public* treatment and observed behavior.

3.4.1 Key Variables of Interest

To test the hypotheses, we generate the following key variables from the experimental data: fairness, beliefs about others' fairness, auction bids, image concerns, and dictator offers in the *public* and *private* treatment (see Table 3.1 for summary statistics).

Table 3.1: Key Variables

	Min	Max	Mean	∅ Share	Std. Dev.
Fairness (DG 1)	0	31	12.4	31%	8.9
Auction bid	0	100	49.0		35.1
Dictator offer (DG 2)	0	100	19.4	19%	20.0
Expected offer (DG 2)	0	100	30.3	30%	19.4
Image	0	1	0.32		0.47

In the initial dictator game DG 1 – played via the strategy method – subjects give on average 31 percent of the surplus to the recipients.

The two most frequent shares given are 50 percent of the surplus (chosen by 41 percent of subjects) and zero (chosen by 27 percent of subjects). This finding also aligns with our assumption about p in the model section. DG 1 is used as a measure of the individual fairness preference θ in the absence of image concerns and role preferences.[7] There is no significant difference in mean offers in DG 1 between the two treatments ($p = 0.204$ t-test, $p = 0.139$ Mann-Whitney-U test, MWU), showing that participants understood that this decision would not be made public and can therefore serve its purpose as a fairness measure independent of image concerns. There are neither significant age nor gender differences in the offers made in DG 1. The variable *fairness* will comprise these dictator offers with a possible range from 0 to 40.

[7] Recall that the instructions for Parts 2 and 3 were not handed out until after the end of Part 1. Before Part 1, participants were told that later on a task would follow in which their decision may or may not be made visible to others. The lab required this upfront announcement because the treatment slightly reduced the usually guaranteed level of anonymity for other participants. Subjects were allowed to reconsider whether they wanted to stay for the experiment. No subject chose to leave.

The *auction bid* can range from 0 to 100 and is elicited from all subjects in the second-price auction preceding DG 2. Likewise, the *dictator offer* made in DG 2 can range from 0 to 100, but is only observed for the subjects ending up in the role of the dictator in DG 2.[8]

In addition to subjects' own actions, we elicit first-order beliefs through a standard incentivized mechanism in which subjects try to predict others' actual behavior. As discussed in Section 3, the first-order belief about the dictator game DG 2 matters for the preceding auction. The variable *expected offer* represents the amount someone expects to receive from another dictator in DG 2.

About one-third of subjects in the *public* treatment reported (on a five-point Likert-scale) that they were strongly or very strongly affected in their decision by the public announcement. For those subjects, we construct the binary variable *image* as a measure of image sensitivity μ, which is 1 for subjects with a strong or very strong effect and 0 for all other subjects (including all participants in the *private* treatment). As such, it represents the subset of subjects in the *public* treatment who self-identify as especially image-concerned.

3.4.2 Protocol and Summary Statistics

The experiment was programmed using the software zTree (Fischbacher, 2007). Six sessions were conducted at the University of Hamburg in October 2020[9]. 162 subjects – recruited via hroot (Bock et al., 2014) – participated in one of the two treatments each (between-subjects design). 82 subjects participated in the *public* treatment and 80 in the *private* (control) treatment. All sessions lasted for less than one hour.

Upon arrival at the lab, subjects were placed anonymously in computer cubicles and received the instructions for Part 1 of the experiment (DG 1). The instructions for Parts 2 and 3 – the auction and DG 2 – were handed out when Part 1 was completed. After reading the instructions, all subjects answered a set of control questions. At the end of each session, subjects completed a non-incentivized questionnaire on socio-demographic characteristics (e.g., age, gender, and field of study). Table 3.2 summarizes the most essential subject pool characteristics.

[8]In both treatments, dictator offers are lower in DG 2 compared to DG 1, suggesting the experimental design did not eliminate entitlement effects. This is in line with the results found by Kassas and Palma (2019), who show that entitlement effects can arise even with random role assignment. However, the role assignment mechanism is identical between treatments and should not bias the treatment effect.

[9]The experimental protocol ensured adherence to the local regulations to limit the spread of Covid-19. Among other measures, subjects and staff were required to wear face masks.

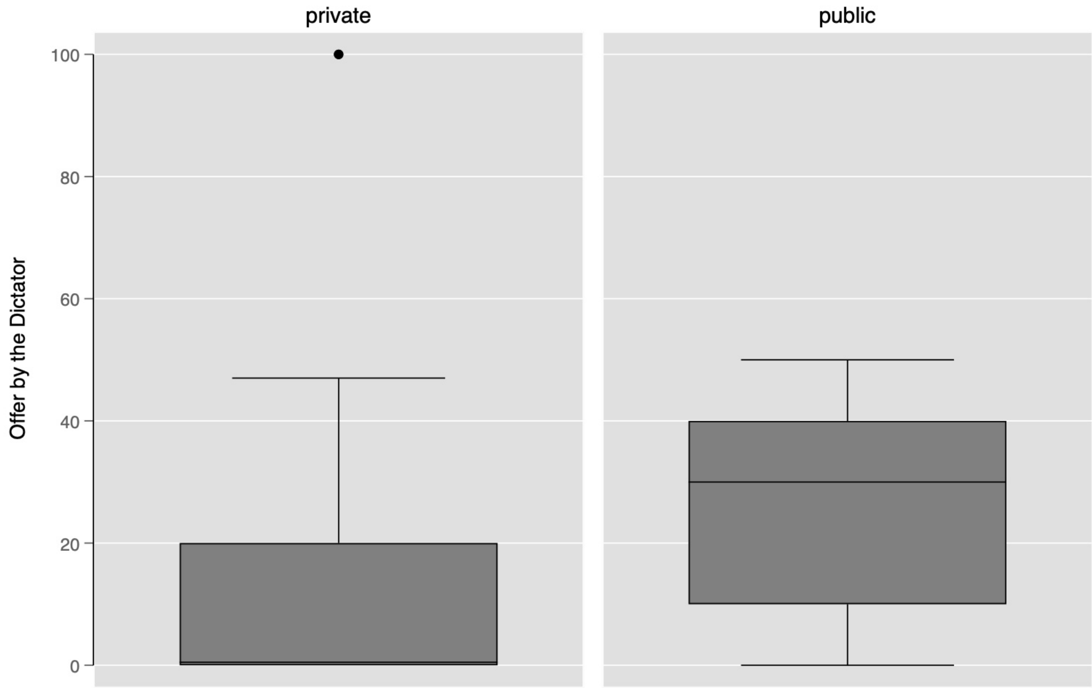
Table 3.2: Summary Statistics

Total number of subjects	162			
Participants in *private* treatment	80			
Participants in *public* treatment	82			
	Min	Max	Mean	Std. Dev.
Age	19	46	26.5	4.9
Payout (in Euros)	5.1	14.3	9.1	1.9
	Share			
Female	69%			
Economics/business student	27%			

3.4.3 Dictator Game (DG 2)

We start the analysis with DG 2, the dictator game played after the auction. The dictator offers observed in DG 2 vary considerably between treatments (see Figure 3.3). The average dictator shares 14 points in the *private* and 24 in the *public* treatment ($p = 0.025$, t-test; $p = 0.005$, MWU). The distribution of offers is significantly different between the treatments ($p = 0.030$, combined Kolmogorov-Smirnov test). Exactly half of all dictators in the *private* treatment share nothing with their opponent, compared to only 22 percent in the *public* treatment. Vice versa, no subject in the *private* treatment shared 50 percent of the pie, compared to 15 percent in the *public* treatment ($p = 0.014$, chi-squared test).

Figure 3.3: Dictator Offers DG 2



Result 1. *The distribution of dictator offers in the public treatment first-order stochastically dominates the distribution in the private treatment. Average dictator offers are higher, and the equal split is more prevalent in the public treatment.*

Result 1 is in line with hypothesis 1 and results from the closest relevant study by Andreoni and Bernheim (2009).

To further investigate how the *public* treatment, fairness, and image concerns influence dictator behavior we continue with a multivariate analysis. Consistent with the simple mean comparison, the *public* treatment has a positive and significant effect on the dictator offers. Table 3.3 shows estimates for a regression analysis of the amount offered by dictators in DG 2. To account for the substantial number of zero offers, we use left-censored Tobit regressions in addition to OLS regressions as a consistency check.

All specifications show not only a highly significant positive effect of the *public* treatment, but also regarding individual fairness. Column one shows a one-to-one relationship between individual fairness, as measured by the offer made in DG 1, and the offer made in DG 2. The evidence regarding an interaction effect between the treatment and fairness is less robust. The Tobit model suggests the treatment effect to vary for different levels of fairness. The treatment effect is most considerable when fairness equals zero. As an individual's fairness increases, the treatment effect decreases. Highly fair-minded individuals seem less affected by image-related stimulation. This is in line with the theoretical considerations about which individuals change their behavior

Table 3.3: Dictator Offers

Dependent Variable: Dictator Offer in DG 2				
	(1)	(2)	(3)	(4)
	OLS	OLS	Tobit	Tobit
Public	15.15***	7.96*	36.86***	14.99***
	(2.78)	(3.27)	(8.84)	(5.03)
Fairness	1.08***	1.16***	2.23***	1.91***
	(0.17)	(0.13)	(0.34)	(0.25)
Public × Fairness	-0.31		-1.29***	
	(0.27)		(0.42)	
Image		23.48**		36.51***
		(5.95)		(7.78)
Image × Fairness		-1.26		-2.18**
		(0.78)		(0.98)
Constant	1.41	0.57	-24.91***	-19.49***
	(1.29)	(2.01)	(7.99)	(5.95)
N	81	81	81	81
R^2	0.25	0.33		
Pseudo R^2			0.06	0.07

Note: *Public* is a dummy variable for the treatment; *fairness* is the offer made in DG 1 (0 to 40); *image* is a dummy variable for self-reported treatment influence. Clustered standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

between treatments: Independent of the treatment, fair-minded individuals choose the high offer, while individuals who are not fair-minded, but image concerned, increase their offer in the *public* treatment.

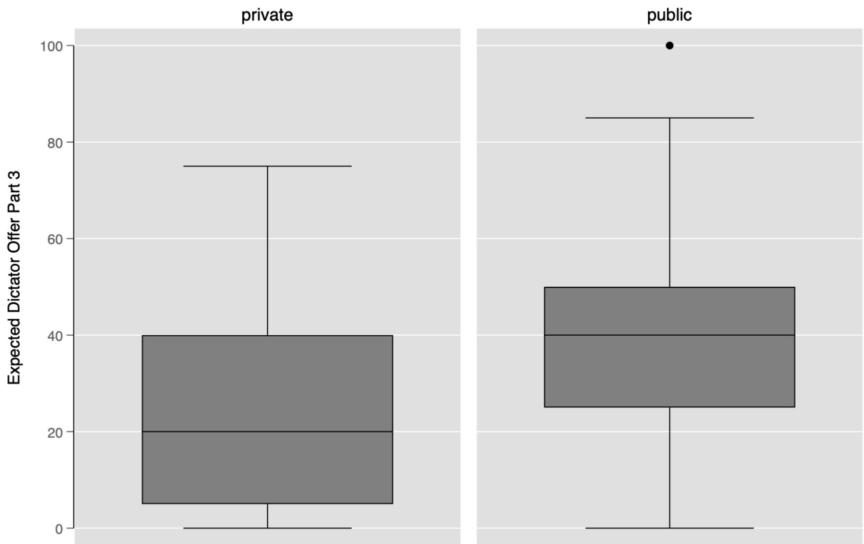
Column two includes the variable image. In comparison to column one, the effect of fairness remains robust, while image sensitivity absorbs a large part of the treatment effect. Having a high image sensitivity increases average offers by 23 points. However, even subjects who claim not to have been strongly affected by the treatment are significantly more generous with the public announcement.[10] The interaction between fairness and image is only significant in the

[10]This could be explained by subjects misreporting their actual image sensitivity, which might be driven by image concerns towards the experimenter or simply by a lack of self-reflection. Alternatively, the auction as a selection mechanism could explain higher offers in the *public* treatment. In section 3.4.4, we show that the average auction bid decreased in the *public* treatment. If certain fairness types drive

Tobit regression in column four. The treatment effect for individuals with high image sensitivity is greatest when they are selfish and decreases with increasing fairness. The impact of fairness itself is canceled out by high image sensitivity, as image concerns drive individuals towards high offers anyway. To summarize, the *public* treatment increases dictator offers, which are in addition positively related to own fairness and image concerns. In coherence with the theory, we see that it is primarily image-concerned individuals who drive the treatment effect.

We conclude the investigation of DG 2 by analyzing how the treatment affects expected dictator offers. Subjects expect on average 61% more points from the dictator in the *public* treatment compared to the *private* (37 vs. 23 points, $p < 0.001$ for t-test and MWU; see Figure 3.4) The distributions of expected offers are significantly different between treatments ($p < 0.001$, combined Kolmogorov-Smirnov test). We see that subjects respond to the treatment and correctly anticipate others' behavior. They expect other subjects to be pro-social (expecting non-zero shares even in the *private* treatment) but also image-concerned and thus responsive to the *public* treatment.

Figure 3.4: Expected Offers in DG 2



3.4.4 Bidding for the Dictator Role

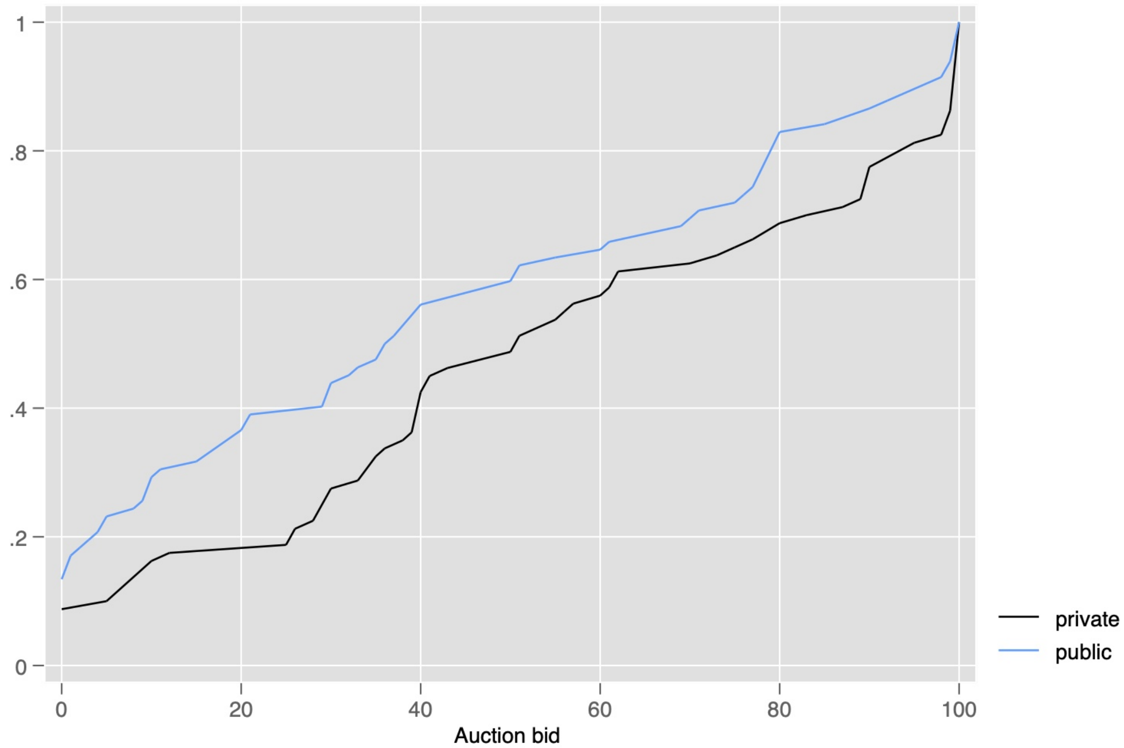
Next, we turn towards the second-price auction, allocating the dictator role in DG 2. The average bid is 43 points in the *public* and 55 in the *private* treatment (t-test $p = 0.035$, MWU $p = 0.024$). Bidding behavior reflects the stated preferences from the post-experiment questionnaire: significantly fewer participants in the *public* treatment prefer the dictator role to the recipient role (t-test: $p = 0.012$, MWU: $p = 0.019$). In the *public* treatment, the mode is

this change in bidding behavior between the treatments, then the set of types that become the dictator potentially differs as well.

bidding 0, whereas the mode in the *private* treatment is 100. Figure 3.5 shows the distribution of auction bids in both treatments. The distribution of bids in the *public* treatment first-order stochastically dominates the distribution in the *private* treatment ($p = 0.038$, combined Kolmogorov-Smirnov test) The evidence supports Hypothesis 2.

Result 2. *The distribution of auction bids in the private treatment first-order dominates the distribution in the public treatment. The average auction bid is higher in the public treatment.*

Figure 3.5: Auction Bids in Part 2



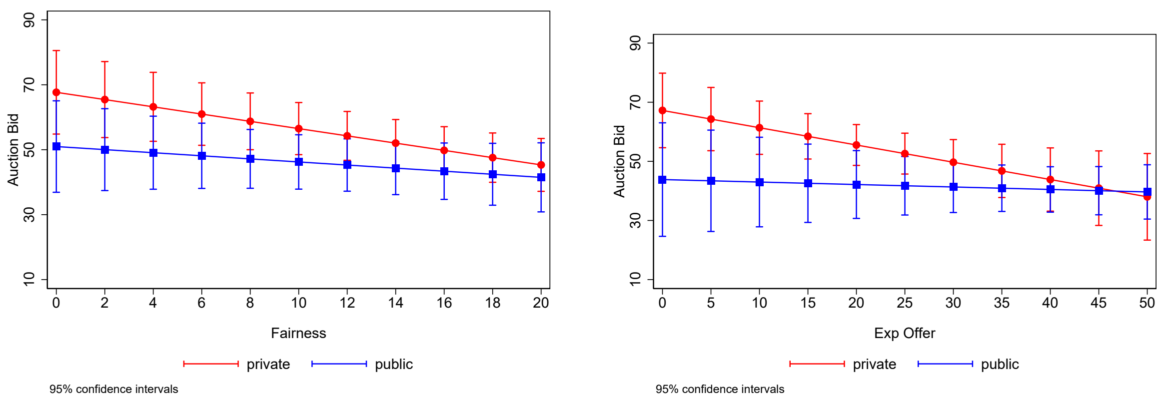
The empirical evidence is in line with the model's predictions in section 3.3.3. The average bid and bidding in general increase across the entire range, with the visibility of the dictator's action – the *public* treatment. Key variables are the fairness preference, image sensitivity, and the expected dictator offer one receives as a recipient.

For the multivariate analysis, we again use Tobit models as consistency checks (Table B1) for the OLS regressions (Table 3.4). Auction bids are limited to values between zero and one hundred. We find clustering at both values and therefore apply two-sided censored Tobit models. Adding interaction terms sheds light on how exactly the image concerns influence behavior. The effect of the *public* treatment on auction bids is not visible in the simple specification in Column 1. But the negative effect of *public* is large and significant when adding interaction terms with fairness and expected offers in Columns 2 and 3. On the one hand, and in line with theory, it is especially individuals with low fairness concerns who are mostly affected by the treatment and bid high in the *private* treatment and low in the *public* treatment. On the other hand, individuals who expect to face rather selfish individuals have a stronger incentive to bid high in the *private* treatment and low in the *public*, as they also correctly (Figure 3.3) expect dictator behavior to swiftly (Figure 3.4) shift between treatments. All specifications show that fairness is also significantly negatively correlated with the auction bid. From section 3.3.3 we are aware that fairness-motivated players have little incentive to engage in rent-seeking in the auction. In addition, expecting a larger offer from the dictator decreases one's auction bid. In terms of our model, expecting many fair-minded individuals, namely a high p, excluded players from choosing a high bid.

Columns 4 through 6 analyze the treatment effect for the subset of image-sensitive subjects within the *public* treatment. We observe a remarkably low willingness to become a dictator, as reflected in the large, negative effect of image on the auction bid: high image concerns decrease the average bid by 29.6-54.4 points, depending on the interacted variable. In Column 5 we interact image with fairness. Here, we see that the effect of image sensitivity depends on the fairness level. Individuals with low fairness preference bid way higher when they are sensitive to their image. As fairness preferences increase, the effect of image sensitivity disappears. In column (6) we find a similar dynamic with the expected offer. Individuals with low expectations bid higher when they are image sensitive than their counterparts. As those expectations increase, the gap between the two groups narrows.

Figure 3.6 shows how the effect of the treatment depends on the level of one's own fairness preference and expected offers. For both variables, the difference between *public* and *private* is driven by the players who are of low fairness and those who expect others to be unfair. Auction bids are much higher in the *private* treatment in those cases. The higher the values of both variables, the more similar the bids are between treatments.

Figure 3.6: Predicted Auction Bids



Result 3. *Auction bids, and hence the willingness to become a dictator, are lower in the public treatment. Fairness, image sensitivity, and expected offers are significant channels of transmission for this effect.*

Result 3 summarizes the findings, which support Hypothesis 2. Both hypotheses cannot be rejected, and the dynamic described in the previous section 3.3.3 on theoretical considerations is supported by the evidence.

3.4.5 Robustness Checks

In addition to the main experiment discussed so far, a modified experiment was run to assess the robustness of the main results regarding specific design choices related to image concerns. First, our results may be driven by factors other than social image concerns. For example, Buser and Yuan (2022) show that public speaking aversion makes subjects willing to pay to avoid giving a public presentation.[11] While subjects in our experiment are not required to speak in front of the other participants, an aversion to personal public appearances may still lower the willingness to pay for the dictator role in the *public* treatment. To control for this effect, we used a modified experimental design for the *public* treatment, in which image concerns are stimulated by showing subjects pictures of the dictators at the end of the session, as used in Andreoni and Petrie (2004). All other aspects of the games were kept as in our original experiment. In the instructions right before the auction, we described the new publication process of the photos, including the dictator's chosen contribution and an identification number.

Second, the previous section identified heterogeneity in image sensitivity as a significant factor explaining the treatment effect. However, image sensitivity is difficult to measure objectively. The variable *image* identified the subjects who were strongly affected by the public announcement. To elicit image concerns independent of the experimental behavior—to limit endogeneity issues—we include another survey measure of image sensitivity in the additional sessions of the *private* and the new *public* treatment. We use a 10-item short form of the Martin-Larsen Approval Motivation Scale (Martin, 1984), which has been found to be a robust measure for social image concerns (Leary et al., 2015).

A total of 144 subjects participated in three sessions with the new *public* treatment and three sessions with the *private* treatment. The modified image sensitivity items were included in the questionnaire for all additional sessions. The sessions were held at the University of Hamburg in February 2023.

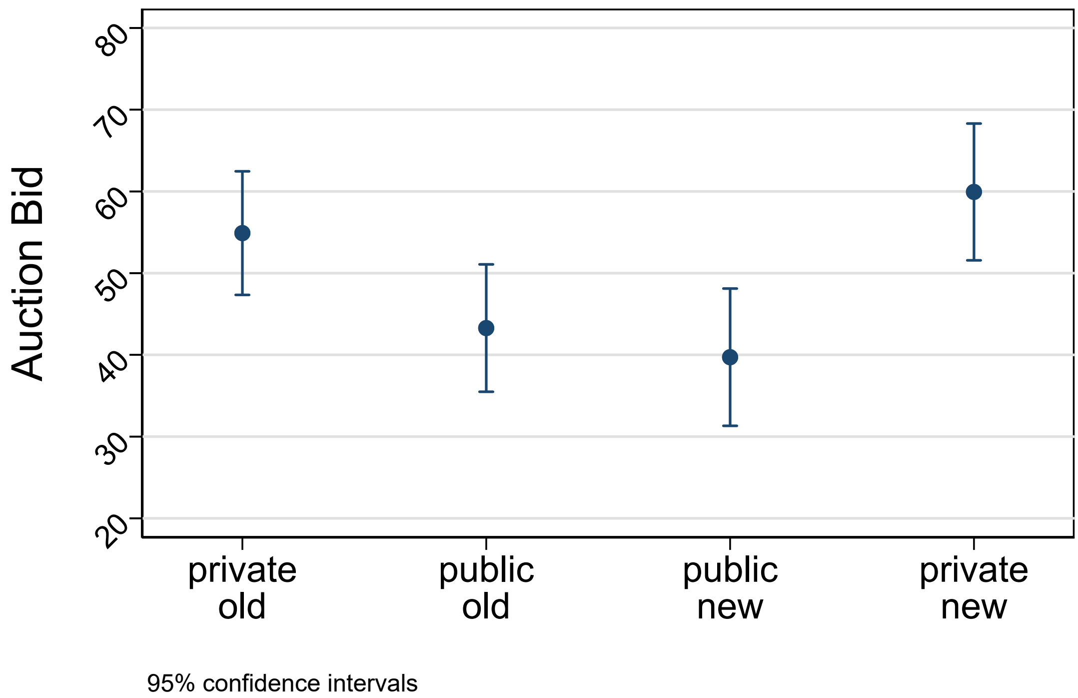
Bidding for the Dictator Role in the Photo Treatment

How do role preferences for the dictator game behave when image concerns are activated via the photo treatment? We find that the decrease in willingness to pay for the dictator role is large when subjects' photos are made public (see Figure 3.7). The *photo* treatment decreases the average auction bid by 34 % compared to the *private* treatment ($p = 0.001$, t-test). The difference in auction bids between the original *public* treatment and the *photo* treatment is not significant ($p = 0.536$, t-test). The specific way in which image concerns are activated does not seem to matter much. Still, the public announcement itself – in person or via the photo – has a robust influence on behavior.

[11]We thank an anonymous referee for this suggestion.

Our other key variables are also robust to the design change. Fairness preferences as measured by dictator offers made in DG 1 are not significantly different between the old *public* and *photo* treatment ($p = 0.412$, t-test) or the *photo* and *private* treatment ($p = 0.645$, t-test). The expected dictator offers in DG 2 are not significantly different between the old *public* and *photo* treatment either ($p = 0.747$, t-test). We can thus conclude that our main results are not driven by an aversion to standing out in public *in person*, but by general image concerns.

Figure 3.7: Comparing Auction Bids between Treatments



Multivariate analysis, summarized in tables B2 and B3, yields similar results in the photo procedure and in the original method.

Measuring Image Sensitivity

Regarding the modified measure for image concerns, a Kolmogorov-Smirnov test does not reject the null hypothesis that the distribution of image concerns is the same across the two treatment conditions ($p = 1.000$). Unsurprisingly, as treatment assignment was random, we should expect equal shares of subjects with high and low image concerns in both conditions. The lower auction bids observed in the *public* treatment, therefore, cannot be attributed to selection effects.

Now, does the new measure for image concerns possess any explanatory power for the different auction bids? In table B4 in the appendix, we present results of a linear regression model of auction bids on fairness, the expected offer, and a dummy encoding high levels in the new image index. Additionally, we included Tobit estimates as a consistency check. We find that neither the high-image dummy alone ($p = 0.607$) nor the interaction between the high-image dummy and the *public* treatment ($p = 0.901$ and $p = 0.752$) has a significant effect on bids.

Ultimately, the refined measure of image concerns has low explanatory power for role preferences. Collecting data on people's social image concerns may not be feasible through direct observation, as this is precisely the crux of those concerns. Once you reveal that social image concerns drive your virtuous behavior, it misses its purpose. Instead, our model presents an argument for why our experimental method can excavate the effect of social image concerns on bargaining without relying on our subjects' honest and reflective statements.

3.5 Discussion and Conclusion

We study role preferences in bargaining in relation to image concerns in a dictator game experiment. Subjects are matched in pairs and participate in a second-price auction. Winning the auction substantially increases the chances of becoming the dictator in the subsequent dictator game. The experimental design systematically varies the visibility of dictators' sharing decisions. In the *private* treatment, dictators' identities are never revealed to recipients, and no information about the amounts they share is provided. In the *public* treatment, at the end of the session, participants gather in the laboratory hallway, and all dictators' decisions are publicly announced.

Previous studies on publicizing dictator decisions showed ambiguous effects on dictator behavior (Hoffman et al., 1994; Bohnet and Frey, 1999a; Dufwenberg and Muren, 2006; Andreoni and Bernheim, 2009). Not only do we shed more light on this empirical controversy, but we also investigate an additional, previously omitted aspect of the otherwise well-studied dictator game: how much subjects prefer the dictator role and how this preference changes with publicity. By applying the second-price auction, we can quantify this role preference. While the dictator holds all the bargaining power in the interaction, she is also forced to reveal *private* information, e.g. about her fairness type, by making the sharing decision. In a behavioral game-theoretic model, we argue that this information revelation entails a psychological cost and is therefore traded off against the monetary gains from the dictator position. Our theoretical argument is strongly inspired by Andreoni and Bernheim (2009)'s signaling model.

We find evidence that our treatment affected subjects' decision-making at both stages of the experiment: the auction and the dictator game. In the dictator game, the *public* treatment increases the average offer by 10 points, and the entire offer distribution is significantly shifted to the right. In a regression analysis, we find that someone's own fairness and image sensitivity indeed matter for generosity in the dictator game. Next to the actual behavior of dictators, we elicit subjects' first-order beliefs of their opponents' behavior as dictators. We find that with the *public* treatment, the distribution of expected offers significantly shifts to higher values. Subjects broadly seem to be convinced of an effect of the *public* treatment on their opponents' behavior. Finally, we find that the average auction bid significantly decreases. The bid distributions significantly differ and shift to lower values in the treatment. Regression analysis suggests that self-proclaimed image concerns have a first-order effect over image concerns. We find a strong negative correlation between fairness, image sensitivity, and auction bids. More specifically, the treatment effect is mainly driven by highly image-sensitive individuals. Furthermore, we find heterogeneity in the impact of fairness. It turns out that the effect peaks for selfish individuals and decreases in fairness. We conclude that especially selfish individuals suffer image costs from publicity and therefore adjust their behavior to avoid informal punishment. Altogether, this decreases their valuation of the dictator role and hence reduces auction bids.

The relevance of the results beyond economic experiments is that increased transparency significantly influences prosocial behavior. Players who care about how others perceive them are motivated to share more of a surplus than they would do otherwise. Instead of benefiting from the first-mover position by anchoring and therefore gaining bargaining power, they are aware of the adverse signaling properties of the first-mover position. These insights matter for bargaining situations of very different kinds: wage negotiations (Rosenfeld and Denice, 2015), corporate social responsibility (Lee and Kohler, 2010), plea bargaining in criminal law (Schneider and Alkon, 2019), or international diplomacy (Stasavage, 2004), to name only a few. Most articles on the topic argue that transparency is beneficial because it reduces information asymmetries. Our results stress another aspect: transparency can reduce inequalities by directly taming the behavior of the more powerful bargaining party. However, we show that the effect is strongest when actors care about how they are perceived. In general this mechanism is at work in any open market economy, for example, if customers adjust their purchasing decisions in reaction to how the public perceives a company.

One aspect of the experimental design worth discussing is that winning the auction increases someone's chances in the following lottery to 90%. This complicates understanding the task and calculating the optimal bidding strategy for our subjects. Bidding on a lottery ticket inevitably invokes highly subjective considerations, including risk preferences and perceptions of probabilities. Nevertheless, we prefer this procedure over an experiment without a lottery, because we want to counter the possible impact of entitlement effects. If it were commonly known that the subjects with the higher bids "bought" the dictator role with certainty, this changes the context of the dictator game substantially. The entitlement effect is well-researched, and our design is therefore better suited for detecting insights about role preferences in the "regular" dictator game.

One aspect of the identification strategy that could be further developed in future studies on the topic is the measurement of image concerns. To create the measure of image sensitivity, we draw on the stated preferences from the questionnaire. The alternative measure used in the robustness check did not corroborate the anticipated effect, underscoring the measure's sensitivity. While there is no (monetary) incentive to misreport image concerns, an objective measure would be more optimal here. Most studies which investigate the effect of social image concerns on behavior as a treatment make the assumption that all subjects are affected equally. Our evidence suggests that this is not necessarily the case. But we are not aware of an (economic) paper providing insights into reliably identifying subjects who can ex ante be expected to be more affected by such a treatment. An interdisciplinary approach seems promising here.

Bibliography

Alevy, J. E., Jeffries, F. L., and Lu, Y. (2014). Gender-and frame-specific audience effects in dictator games. *Economics Letters*, 122(1):50–54.

Andersson, P. A., Erlandsson, A., Västfjäll, D., and Tinghög, G. (2020). Prosocial and moral behavior under decision reveal in a public environment. *Journal of Behavioral and Experimental Economics*, 87:101561.

Andreoni, J. and Bernheim, D. (2009). Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636.

Andreoni, J. and Petrie, R. (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *Journal of public Economics*, 88(7-8):1605–1623.

Andreoni, J., Rao, J., and Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy*, 125(3):625–653.

Barmettler, F., Fehr, E., and Zehnder, C. (2012). Big experimenter is watching you! anonymity and prosocial behavior in the laboratory. *Games and Economic Behavior*, 75(1):17–34.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

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

Bohnet, I. and Frey, B. S. (1999a). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, 89(1):335–339.

Bohnet, I. and Frey, B. S. (1999b). The sound of silence in prisoner's dilemma and dictator games. *Journal of economic behavior & organization*, 38(1):43–57.

Bolton, G., Dimant, E., and Schmidt, U. (2021). Observability and social image: On the robustness and fragility of reciprocity. *Journal of Economic Behavior & Organization*, 191:946–964.

Bolton, G. E., Katok, E., and Zwick, R. (1998). Dictator game giving: Rules of fairness versus acts of kindness. *International Journal of Game Theory*, 27(2):269–299.

Broberg, T., Ellingsen, T., and Johannesson, M. (2007). Is generosity involuntary? *Economics Letters*, 94(1):32–37.

Brosig, J., Weimann, J., and Yang, C.-L. (2003). The hot versus cold effect in a simple bargaining experiment. *Experimental Economics*, 6(1):75–90.

Buser, T. and Yuan, H. (2022). Public speaking aversion. *Management Science*.

Camerer, C. F. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Charness, G. and Gneezy, U. (2008). What's in a name? anonymity and social distance in dictator and ultimatum games. *Journal of Economic Behavior & Organization*, 68(1):29–35.

Dal Bó, P., Foster, A., and Putterman, L. (2010). Institutions and behavior: Experimental evidence on the effects of democracy. *American Economic Review*, 100(5):2205–29.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and Human Decision Processes*, 100(2):193–201.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Dreber, A., Ellingsen, T., Johannesson, M., and Rand, D. G. (2013). Do people care about social context? framing effects in dictator games. *Experimental Economics*, 16(3):349–371.

Dufwenberg, M. and Muren, A. (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization*, 61(1):42–49.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4):583–610.

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

Forsythe, R., Horowitz, J., Savin, N., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3):347–369.

Franzen, A. and Pointner, S. (2012). Anonymity in the dictator game revisited. *Journal of Economic Behavior & Organization*, 81(1):74–81.

Frey, B. S. and Bohnet, I. (1995). Institutions affect fairness: Experimental investigations. *Journal of Institutional and Theoretical Economics*, pages 286–303.

Frohlich, N., Oppenheimer, J., and Kurki, A. (2004). Modeling other-regarding preferences and an experimental test. *Public Choice*, 119(1-2):91–117.

Frohlich, N., Oppenheimer, J., and Moore, J. B. (2001). Some doubts about measuring self-interest using dictator experiments: the costs of anonymity. *Journal of Economic Behavior & Organization*, 46(3):271–290.

García-Gallego, A., Georgantzís, N., and Jaramillo-Gutiérrez, A. (2008). Ultimatum salary bargaining with real effort. *Economics Letters*, 98(1):78–83.

Güth, W. and Tietz, R. (1986). Auctioning ultimatum bargaining position. *Current Issues in West German Decision Research, Lang, Frankfurt, Germany*, pages 60–73.

Hoffman, E., McCabe, K., Shachat, K., and Smith, V. (1994). Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*, 7(3):346–380.

Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *American Economic Review*, 86(3):653–660.

Johannesson, M. and Persson, B. (2000). Non-reciprocal altruism in dictator games. *Economics Letters*, 69(2):137–142.

Jones, D. and Linardi, S. (2014). Wallflowers: Experimental evidence of an aversion to standing out. *Management Science*, 60(7):1757–1771.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of Business*, pages S285–S300.

Kassas, B. and Palma, M. A. (2019). Self-serving biases in social norm compliance. *Journal of Economic Behavior & Organization*, 159:388–408.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Leary, M., Jongman-Sereno, K., and Diebels, K. (2015). Measures of concerns with public image and social evaluation. In *Measures of personality and social psychological constructs*, pages 448–473. Elsevier.

Lee, M. and Kohler, J. (2010). Benchmarking and transparency: Incentives for the pharmaceutical industry's corporate social responsibility. *Journal of Business Ethics*, 95(4):641–658.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2):426–432.

Martin, H. (1984). A revised measure of approval motivation and its relationship to social desirability. *Journal of personality assessment*, 48(5):508–519.

Rankin, F. W. (2006). Requests and social distance in dictator games. *Journal of Economic Behavior & Organization*, 60(1):27–36.

Rosenfeld, J. and Denice, P. (2015). The power of transparency: evidence from a british workplace survey. *American Sociological Review*, 80(5):1045–1068.

Schneider, A. K. and Alkon, C. (2019). Bargaining in the dark: The need for transparency and data in plea bargaining. *New Criminal Law Review*, 22(4):434–493.

Shachat, J. and Swarthout, J. T. (2013). Auctioning the right to play ultimatum games and the impact on equilibrium selection. *Games*, 4(4):738–753.

Stasavage, D. (2004). Open-door or closed-door? transparency in domestic and international bargaining. *International Organization*, 58(4):667–703.

Table 3.4: Auction Bids

	OLS Estimates Dependent Variable: Auction Bid					
	(1)	(2)	(3)	(4)	(5)	(6)
Public	-8.491	-16.70*	-23.39**			
	(5.807)	(9.744)	(11.41)			
Fairness	-0.792**	-1.118***	-0.735**	-0.970**	-1.674***	-1.007**
	(0.306)	(0.363)	(0.301)	(0.467)	(0.607)	(0.461)
Exp Offer	-0.323*	-0.307*	-0.584**	0.0529	0.0597	-0.0645
	(0.177)	(0.180)	(0.239)	(0.204)	(0.207)	(0.241)
Public × Fairness		0.643				
		(0.577)				
Public × Exp Offer			0.501			
			(0.326)			
Image				-29.61***	-50.17***	-54.43***
				(8.293)	(11.28)	(16.07)
Image × Fairness					2.030**	
					(0.852)	
Image × Exp Offer						0.615*
						(0.338)
Constant	72.93***	76.88***	78.24***	61.86***	70.66***	66.49***
	(5.888)	(6.336)	(6.628)	(9.414)	(9.893)	(10.26)
N	162	162	162	82	82	82
R^2	0.114	0.120	0.130	0.163	0.217	0.178

Note: *Public* is a dummy variable for the treatment; *fairness* is the offer made in Part 1 (0 to 40); *exp offer* is the first-order belief about others' offers; *image* is a dummy variable for self-reported treatment influence; Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

A Theory Part

A.1 Calculating Posterior Beliefs

There is no equilibrium in which type t_1 chooses o_l, therefore we can neglect $Pr(t_1|o_l)$.
To get the posterior $Pr(t_1|o_h)$, we need to sum up probabilities of all scenarios where the events (o_h, t_1) and (o_h, t_2) occur to find $Pr(o_h, t_1)$ and $Pr(o_h, t_2)$. An event in which player i becomes the dictator and thereby signals its type to the audience is a collection of elementary events that coincide.

$$\begin{aligned} Pr(o_h^i, t_1^i) = & Pr(t_1^i, t_1^j, b_h^i, b_h^j, D_i, o_h^i) + Pr(t_1^i, t_1^j, b_h^i, b_l^j, o_h^i) + Pr(t_1^i, t_1^j, b_l^i, b_l^j, D_i, o_h^i) \\ & + Pr(t_1^i, t_2^j, b_h^i, b_h^j, D_i, o_h^i) + Pr(t_1^i, t_2^j, b_h^i, b_l^j, o_h^i) + Pr(t_1^i, t_2^j, b_l^i, b_l^j, D_i, o_h^i) \end{aligned} \tag{3.12}$$

$$\begin{aligned} & Pr(t_1^i, t_1^j, b_h^i, b_h^j, D_i, o_h^i) = \\ & \quad Pr(o_h^i|t_1^i, t_1^j, b_h^i, b_h^j, D_i) Pr(D_i|t_1^i, t_1^j, b_h^i, b_h^i) Pr(b_h^j|t_1^i, t_1^j, b_h^i) Pr(b_h^i|t_1^i, t_1^j) Pr(t_1^j|t_1^i) Pr(t_1^i) \\ & = \alpha_i(I_3^i) \cdot \frac{1}{2} \cdot \beta_j(I_1^j) \cdot \beta_i(I_1^i) \cdot p \cdot p \end{aligned} \tag{3.13}$$

$$\begin{aligned} & Pr(t_1^i, t_1^j, b_h^i, b_l^j, o_h^i) = \\ & \quad Pr(o_h^i|t_1^i, t_1^j, b_h^i, b_l^j) Pr(b_l^j|t_1^i, t_1^j, b_h^i) Pr(b_h^i|t_1^i, t_1^j) Pr(t_1^j|t_1^i) Pr(t_1^i) \\ & = \alpha_i(I_3^i) \cdot \beta_j(I_1^j) \cdot \beta_i(I_1^i) \cdot p \cdot p \end{aligned} \tag{3.14}$$

$$\begin{aligned} & Pr(t_1^i, t_1^j, b_l^i, b_l^j, D_i, o_h^i) = \\ & \quad Pr(o_h^i|t_1^i, t_1^j, b_l^i, b_l^j, D_i) Pr(D_i|t_1^i, t_1^j, b_l^i, b_l^j) Pr(b_l^j|t_1^i, t_1^j, b_l^i) Pr(b_h^i|t_1^i, t_1^j) Pr(t_1^j|t_1^i) Pr(t_1^i) \\ & = \alpha_i(I_4^i) \cdot \frac{1}{2} \cdot \beta_j(I_1^j) \cdot \beta_i(I_1^i) \cdot p \cdot p \end{aligned} \tag{3.15}$$

$$\begin{aligned} & Pr(t_1^i, t_2^j, b_h^i, b_h^j, D_i, o_h^i) = \\ & \quad Pr(o_h^i|t_1^i, t_2^j, b_h^i, b_h^j, D_i) Pr(D_i|t_1^i, t_2^j, b_h^i, b_h^j) Pr(b_h^j|t_1^i, t_2^j, b_h^i) Pr(b_h^i|t_1^i, t_2^j) Pr(t_2^j|t_1^i) Pr(t_1^i) \\ & = \alpha_i(I_3^i) \cdot \frac{1}{2} \cdot \beta_j(I_2^j) \cdot \beta_i(I_1^i) \cdot (1-p) \cdot p \end{aligned} \tag{3.16}$$

$$\begin{aligned} & Pr(t_1^i, t_2^j, b_h^i, b_l^j, o_h^i) = \\ & \quad Pr(o_h^i|t_1^i, t_2^j, b_h^i, b_l^j) Pr(b_l^j|t_1^i, t_2^j, b_h^i) Pr(b_h^i|t_1^i, t_2^j) Pr(t_2^j|t_1^i) Pr(t_1^i) \\ & = \alpha_i(I_3^i) \cdot \frac{1}{2} \cdot \beta_j(I_2^j) \cdot \beta_i(I_1^i) \cdot (1-p) \cdot p \end{aligned} \tag{3.17}$$

$$\begin{aligned} & Pr(t_1^i, t_2^j, b_l^i, b_l^j, D_i, o_h^i) = \\ & \quad Pr(o_h^i|t_1^i, t_2^j, b_l^i, b_l^j, D_i) Pr(D_i|t_1^i, t_2^j, b_l^i, b_l^j) Pr(b_l^j|t_1^i, t_2^j, b_l^i) Pr(b_l^i|t_1^i, t_2^j) Pr(t_2^j|t_1^i) Pr(t_1^i) \\ & = \alpha_i(I_4^i) \cdot \frac{1}{2} \cdot \beta_j(I_2^j) \cdot \beta_i(I_1^i) \cdot (1-p) \cdot p \end{aligned} \tag{3.18}$$

$$\begin{aligned} Pr(o_h^i, t_2^i) =& Pr(t_2^i, t_1^j, b_h^i, b_h^j, D_i, o_h^i) + Pr(t_2^i, t_1^j, b_h^i, b_l^j, o_h^i) + Pr(t_2^i, t_1^j, b_l^i, b_l^j, D_i, o_h^i) \\ &+ Pr(t_2^i, t_2^j, b_h^i, b_h^j, D_i, o_h^i) + Pr(t_2^i, t_2^j, b_h^i, b_l^j, o_h^i) + Pr(t_2^i, t_2^j, b_l^i, b_l^j, D_i, o_h^i) \end{aligned} \tag{3.19}$$

$$\begin{aligned} &Pr(t_2^i, t_1^j, b_h^i, b_h^j, D_i, o_h^i) = \\ &\quad Pr(o_h^i|t_2^i, t_1^j, b_h^i, b_h^j, D_i)Pr(D_i|t_2^i, t_1^j, b_h^i, b_h^j)Pr(b_h^j|t_2^i, t_1^j, b_h^i)Pr(b_h^i|t_2^i, t_1^j)Pr(t_1^j|t_2^i)Pr(t_2^i) \\ &= \alpha_i(I_5^i) \cdot \frac{1}{2} \cdot \beta_j(I_1^j) \cdot \beta_i(I_2^i) \cdot p \cdot (1-p) \end{aligned} \tag{3.20}$$

$$\begin{aligned} &Pr(t_2^i, t_1^j, b_h^i, b_l^j, o_h^i) = \\ &\quad Pr(o_h^i|t_2^i, t_1^j, b_h^i, b_l^j)Pr(b_l^j|t_2^i, t_1^j, b_h^i)Pr(b_h^i|t_2^i, t_1^j)Pr(t_1^j|t_2^i)Pr(t_2^i) \\ &= \alpha_i(I_5^i) \cdot \frac{1}{2} \cdot \beta_j(I_1^j) \cdot \beta_i(I_2^i) \cdot p \cdot (1-p) \end{aligned} \tag{3.21}$$

$$\begin{aligned} &Pr(t_2^i, t_1^j, b_l^i, b_l^j, D_i, o_h^i) = \\ &\quad Pr(o_h^i|t_2^i, t_1^j, b_l^i, b_l^j, D_i)Pr(D_i|t_2^i, t_1^j, b_l^i, b_l^j)Pr(b_l^j|t_2^i, t_1^j, b_l^i)Pr(b_h^i|t_2^i, t_1^j)Pr(t_1^j|t_2^i)Pr(t_2^i) \\ &= \alpha_i(I_6^i) \cdot \frac{1}{2} \cdot \beta_j(I_1^j) \cdot \beta_i(I_2^i) \cdot p \cdot (1-p) \end{aligned} \tag{3.22}$$

$$\begin{aligned} &Pr(t_2^i, t_2^j, b_h^i, b_h^j, D_i, o_h^i) = \\ &\quad Pr(o_h^i|t_2^i, t_2^2, b_h^i, b_h^j, D_i)Pr(D_i|t_2^i, t_2^2, b_h^i, b_h^j)Pr(b_h^j|t_2^i, t_2^2, b_h^i)Pr(b_h^i|t_2^i, t_2^2)Pr(t_2^2|t_2^i)Pr(t_2^i) \\ &= \alpha_i(I_5^i) \cdot \frac{1}{2} \cdot \beta_j(I_2^j) \cdot \beta_i(I_2^i) \cdot (1-p) \cdot (1-p) \end{aligned} \tag{3.23}$$

$$\begin{aligned} &Pr(t_2^i, t_2^j, b_l^i, b_l^j, D_i, o_h^i) = \\ &\quad Pr(o_h^i|t_2^i, t_2^2, b_h^i, b_l^j)Pr(b_l^j|t_2^i, t_2^2, b_h^i)Pr(b_h^i|t_2^i, t_2^2)Pr(t_2^2|t_2^i)Pr(t_2^i) \\ &= \alpha_i(I_5^i) \cdot \frac{1}{2} \cdot \beta_j(I_2^j) \cdot \beta_i(I_2^i) \cdot (1-p) \cdot (1-p) \end{aligned} \tag{3.24}$$

$$\begin{aligned} &Pr(t_2^i, t_2^j, b_l^i, b_l^j, D_i, o_h^i) = \\ &\quad Pr(o_h^i|t_2^i, t_2^2, b_l^i, b_l^j, D_i)Pr(D_i|t_2^i, t_2^2, b_l^i, b_l^j)Pr(b_l^j|t_2^i, t_2^2, b_l^i)Pr(b_l^i|t_2^i, t_2^2)Pr(t_2^2|t_2^i)Pr(t_2^i) \\ &= \alpha_i(I_6^i) \cdot \frac{1}{2} \cdot \beta_j(I_2^j) \cdot \beta_i(I_2^i) \cdot (1-p) \cdot (1-p) \end{aligned} \tag{3.25}$$

A.2 Analysis

Proposition 1:

$$U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_1) = -(o^{eq} - o_h) = 0 \tag{3.26}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_1) = -(o^{eq} - o_l) = -(\frac{1}{2} - o_l) \tag{3.27}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_1) \geq & U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_1) \\ o_l \leq & \frac{1}{2} \end{aligned} \tag{3.28}$$

By definition $o_h > o_l$ and therefore the t_1 type will never choose o_l.

Proposition 2: Type $t_2 = (0, \frac{2}{3})$ is partially interested in financial gains and their image. When deciding between o_h and o_l in information sets I^i_κ, with $\kappa = 5, 6$, they compare utilities and prefer o_h over o_l under the following condition.

$$U_i^D(\alpha_i(I^i_\kappa) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) - \frac{2}{3}\sigma(1 - E[\theta|o_h]) \tag{3.29}$$

$$U_i^D(\alpha_i(I^i_\kappa) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma(1 - E[\theta|o_l]) \tag{3.30}$$

$$\begin{aligned} U_i^D(\alpha_i(I^i_\kappa) = 1; \sigma, t_2) \geq & U_i^D(\alpha_i(I^i_\kappa) = 0; \sigma, t_2) \\ o_h - o_l \leq & 2\sigma(E[\theta|o_h] - E[\theta|o_l]) \\ o_l \geq & \frac{1}{2} - 2\sigma(E[\theta|o_h] - E[\theta|o_l]) \end{aligned} \tag{3.31}$$

Proposition 3: Only in the *private* treatment does a separating equilibrium exist. A t_2-type dictator only chooses o_l over o_h, if the following condition is fulfilled.

$$U_i^D(\alpha_i(I^i_\kappa) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) - \frac{2}{3}\sigma(1 - 1) = \frac{1}{3}\frac{1}{2} = \frac{1}{6} \tag{3.32}$$

$$U_i^D(\alpha_i(I^i_\kappa) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma(1 - 0) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma = \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \tag{3.33}$$

$$\begin{aligned} U_i^D(\alpha_i(I^i_\kappa) = 0; \sigma, t_2) > U_i^D(\alpha_i(I^i_\kappa) = 1; \sigma, t_2) \\ \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma > \frac{1}{6} \\ o_l < \frac{1}{2} - 2\sigma \\ o_l < -\frac{3}{2} \quad (\textit{public } \sigma = 1) \\ o_l < \frac{3}{10} \quad (\textit{private } \sigma = \frac{1}{10}) \end{aligned} \tag{3.34}$$

Separating Equilibrium Candidates

$\mathbf{s_i = (1, 1, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I^i_1) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I^i_3); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I^j_3), \sigma, t_1)\Big)\Big] \\ & + (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I^i_3); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I^j_5), \sigma, t_1))\Big)\Big] \\ =& 1 + \frac{1}{2}p(-bh) + \frac{1}{2}(1-p)(-bh - (\frac{1}{2} - o_l)) \\ =& \frac{1}{4}\Big(3 - 2b_h + p + 2o_l - 2po_l\Big) \end{aligned} \tag{3.35}$$

$$\begin{aligned} U_i^A(\beta_i(I^i_1) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I^j_3), \sigma, t_1)\Big)\Big] + (1-p))\Big[\Big(U_i^R(\alpha_j(I^j_5), \sigma, t_1)\Big)\Big] \\ =& 1 + (1-p)(-(\frac{1}{2} - o_l)) \\ =& \frac{1}{2}\Big(1 + p + 2o_l - 2po_l\Big) \end{aligned} \tag{3.36}$$

$$\begin{aligned}U_i^A(\beta_i(I_1^i)=1,s_j;\sigma,t_1) \geq& U_i^A(\beta_i(I_1^i)=0,s_j;\sigma,t_1)\\ \frac{1}{4}\Big(3-2b_h+p+2o_l-2po_l\Big) \geq& \frac{1}{2}\Big(1+p+2o_l-2po_l\Big)\\ p \leq& \frac{1-2b_h-2o_l}{1-2o_l}\end{aligned} \tag{3.37}$$

In the *private* treatment for $b_h < \frac{1}{2} - o_l$ the fraction is positive. For $p < \frac{1-2b_h-2o_l}{1-2o_l}$, t_1 chooses b_h.

Auction Decision: t_2 chooses b_h, if

$$\begin{aligned}U_i^A(\beta_i(I_2^i)=1,s_j;\sigma,t_2) =& 1+p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i);\sigma,t_2)-b_h)+\frac{1}{2}(U_i^R(\alpha_j(I_3^j),\sigma,t_2)\Big)\Big]\\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i);\sigma,t_2)-b_h)+\frac{1}{2}U_i^R(\alpha_j(I_5^j),\sigma,t_2)\Big)\Big]\\ =& 1+\frac{1}{2}p\Big(\frac{1}{3}(1-o_l)-\frac{2}{3}\sigma-b_h+\frac{1}{3}(\frac{1}{2})\Big)\\ &+\frac{1}{2}(1-p)\Big(\frac{1}{3}(1-o_l)-\frac{2}{3}\sigma-b_h+\frac{1}{3}(o_l)\Big)\\ =& \frac{1}{12}\Big(14-6b_h+p-2po_l-4\sigma]\Big)\end{aligned} \tag{3.38}$$

$$\begin{aligned}U_i^A(\beta_i(I_2^i)=0,s_j;\sigma,t_2) =& 1+p\Big[\Big(U_i^R(\alpha_j(I_3^j),\sigma,t_2)\Big)\Big]+(1-p))\Big[\Big(U_i^R(\alpha_j(I_5^j),\sigma,t_2)\Big)\Big]\\ =& 1+p\Big(\frac{1}{3}\frac{1}{2}\Big)+(1-p)\Big(\frac{1}{3}o_l\Big)\\ =& \frac{1}{6}\Big(6+p+2xl-2pxl\Big)\end{aligned} \tag{3.39}$$

$$\begin{aligned}U_i^A(\beta_i(I_2^i)=1,s_j;\sigma,t_2) >& U_i^A(\beta_i(I_2^i)=0,s_j;\sigma,t_2)\\ \frac{1}{12}\Big(14-6b_h+p-2po_l-4\sigma\Big) >& \frac{1}{6}\Big(6+p+2xl-2pxl\Big)\\ p <& \frac{8-30b_h-20o_l}{5-10o_l} \quad (\textit{private } \sigma=\frac{1}{10})\end{aligned} \tag{3.40}$$

In the *private* treatment for $b_h < \frac{1}{30(3-10o_l)}$ the fraction is greater than 1 and for any p, t_1 chooses b_h. For $\frac{1}{30}(3-10o_l) < b_h < \frac{4}{15} - \frac{2}{3}o_l$ the fraction is positive. For $p < \frac{8-30b_h-20o_l}{5-10o_l}$, t_2 chooses b_h.

$\mathbf{s_i = (1,0,1,1,0,0)}$

Auction Decision: t_1 chooses b_h, if

$$\begin{aligned}U_i^A(\beta_i(I_1^i)=1,s_j;\sigma,t_1) =& 1+p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i);\sigma,t_1)-b_h)+\frac{1}{2}(U_i^R(\alpha_j(I_3^j),\sigma,t_1)\Big)\\ &+(1-p)\Big[\Big(U_i^D(\alpha_i(I_3^i);\sigma,t_1)\Big)\Big]\\ =& 1+\frac{1}{2}p\Big(-b_h-\frac{1}{2}+o_l\Big)\\ =& 1-\frac{1}{4}p(1+2b_h-2o_l)\end{aligned} \tag{3.41}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_6^j), \sigma, t_1)\Big)\Big] \\ =& 1 + \frac{1}{2}(1-p)(-\frac{1}{2} + o_l) \\ =& \frac{1}{4}(3 + p + 2o_l - 2po_l) \end{aligned} \tag{3.42}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \\ 1 - \frac{1}{4}p(1 + 2b_h - 2o_l) \geq& \frac{1}{4}(3 + p + 2o_l - 2po_l) \\ p \leq& \frac{1 - 2o_l}{2 + 2b_h - 4o_l} > 0 \end{aligned} \tag{3.43}$$

In the *private* treatment for any $b_h > 0$, the fraction is positive. For $p \leq \frac{1-2o_l}{2+2b_h-4o_l}$, t_1 chooses b_h.

Auction Decision: t_2 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_2)\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ =& 1 + \frac{1}{2}p(\frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma - b_h + \frac{1}{6}) + (1-p)(\frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma) \\ =& \frac{1}{12}(-4(-4 + 2s + o_l) + p(-1 - 6b_h + 4s + 2o_l)) \end{aligned} \tag{3.44}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_2)\Big) + \Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_6^j), \sigma, t_2))\Big)\Big] \\ =& 1 + p(\frac{1}{6}) + \frac{1}{2}(1-p)(\frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma + \frac{1}{3}o_l) \\ =& \frac{1}{6}(7 + 2(-1+p)s) \end{aligned} \tag{3.45}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \geq& U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \\ \frac{1}{6}(7 + 2(-1+p)\sigma) \geq& \frac{1}{12}(-4(-4 + 2\sigma + o_l) + p(-1 - 6b_h + 4\sigma + 2o_l)) \\ p \geq& \frac{8 - 20o_l}{5 + 30b_h - 10o_l} \qquad (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.46}$$

In the *private* treatment for $bh > \frac{1}{30}(3 - 10o_l)$, the fraction is positive and smaller 1. For $p \geq \frac{8-20o_l}{5+30b_h-10o_l}$, t_2 chooses b_l

$\mathbf{s_i = (0, 1, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_l, if

$$\begin{aligned}U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big)\Big] \\ =& 1 + \frac{1}{2}(1-p)(-b_h - \frac{1}{2} + o_l) \\ =& 1 + \frac{1}{4}(-1+p)(1+2b_h - 2o_l)\end{aligned} \tag{3.47}$$

$$\begin{aligned}U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big] \\ &+(1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_1)\Big)\Big] \\ =& 1 + (1-p)(-\frac{1}{2} + o_l) \\ =& \frac{1}{2}(1 + p + 2o_l - 2po_l)\end{aligned} \tag{3.48}$$

$$\begin{aligned}U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \\ \frac{1}{2}(1 + p + 2o_l - 2po_l) \geq& 1 + \frac{1}{4}(-1+p)(1+2b_h - 2o_l) \\ p&\begin{cases} \geq 1 & , if \quad b_h \leq \frac{1}{2} - o_l \\ < 1 & , if \quad b_h > \frac{1}{2} - o_l \end{cases}\end{aligned} \tag{3.49}$$

In the *private* treatment for $b_h > \frac{1}{2} - o_l$ and any p, t_1 chooses b_l.

Auction Decision: t_2 chooses b_h, if

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\ =& 1 + p(\frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma) + \frac{1}{2}(1-p)(\frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma - b_h + \frac{1}{3}o_l) \\ =& \frac{1}{6}(7 + 3b_h(-1+p) + p - 2\sigma - 2p\sigma - 2po_l)\end{aligned} \tag{3.50}$$

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_4^j), \sigma, t_2)\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\ =& 1 + \frac{1}{2}p(\frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma + \frac{1}{6}) + (1-p)(\frac{1}{3}o_l) \\ =& \frac{1}{12}(p(3 - 4\sigma - 6o_l) + 4(3 + o_l))\end{aligned} \tag{3.51}$$

$$\begin{aligned}&U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \geq U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \\ &\frac{1}{6}(7 + 3b_h(-1+p) + p - 2\sigma - 2p\sigma - 2po_l) \geq \frac{1}{12}(p(3 - 4\sigma - 6o_l) + 4(3 + o_l)) \\ &for\ o_l < \frac{3}{10} \qquad (private\ \sigma = \frac{1}{10}) \\ &p\begin{cases} \leq \frac{8-30b_h-20o_l}{5-30b_h-10o_l} & , if \quad b_h \leq \frac{1}{6}(1-2o_l) \\ > \frac{8-30b_h-20o_l}{5-30b_h-10o_l} & , if \quad b_h > \frac{1}{6}(1-2o_l) \end{cases}\end{aligned} \tag{3.52}$$

In the *private* treatment for $o_l \leq \frac{3}{10}$ and $b_h \leq \frac{1}{6}(1 - 2o_l)$, the fraction is positive and larger 1.

For any p, t_2 chooses b_h. For $b_h > \frac{1}{6}(1-2o_l)$ and $p > \frac{8-30b_h-20o_l}{5-30b_h-10o_l}$, t_2 chooses b_h.

$\mathbf{s_i = (0, 0, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ =& 1 \end{aligned} \tag{3.53}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_6^j), \sigma, t_1)\Big)\Big] \\ =& 1 + \frac{1}{2}(1-p)(-\frac{1}{2} + o_l) \\ =& \frac{1}{4}(3 + p + 2o_l - 2po_l) \end{aligned} \tag{3.54}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \\ \frac{1}{4}(3 + p + 2o_l - 2po_l) \geq& 1 \\ p \geq& 1 \end{aligned} \tag{3.55}$$

In the *private* treatment for $o_l < \frac{3}{10}$, p would need to be greater than 1, therefore t_1 will not choose b_h.

Pooling Equilibrium Candidates

$\mathbf{s_i = (1, 1, 1, 1, 1, 1)}$

Audience's belief

$$\begin{aligned} Pr(o_h, t_1) =& pp\frac{1}{2} + p(1-p)\frac{1}{2} = \frac{1}{2}p \\ Pr(o_h, t_2) =& (1-p)p\frac{1}{2} + (1-p)(1-p)\frac{1}{2} = (1-p)\frac{1}{2} \\ Pr(t_1|o_h) =& \frac{\frac{1}{2}p}{\frac{1}{2}p + (1-p)\frac{1}{2}} = p \\ E[t_1|o_h] =& p \end{aligned} \tag{3.56}$$

Let $\kappa = 5, 6$. Dictator Decision: t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) - \frac{2}{3}\sigma(1-p) = \frac{1}{3}\frac{1}{2} - \frac{2}{3}\sigma(1-p) = \frac{1}{6}(1 + 4(-1+p)\sigma) \tag{3.57}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma(1-0) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma = \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \tag{3.58}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) &> U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) \\ \frac{1}{6}(1 + 4(-1+p)\sigma) &> \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \\ p &\geq \frac{1}{4}(1 - 2o_l) \quad (\textit{public } \sigma = 1) \\ p &\geq \frac{1}{2}(5 - 10o_l) \quad (\textit{private } \sigma = \frac{1}{10}) \end{aligned} \tag{3.59}$$

In the *public* treatment $p > \frac{1}{4}(1 - 2o_l)$. In the *private* treatment $o_l > \frac{3}{10}$ and $p > \frac{1}{2}(5 - 10o_l)$.

Auction Decision t_1 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\ &+(1-p))\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big)\Big] \\ =& 1 + \frac{1}{2}p(-b_h) + \frac{1}{2}(1-p)(-b_h) \\ =& 1 - \frac{1}{2}b_h \end{aligned} \tag{3.60}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_1)\Big)\Big] \\ =& 1 \end{aligned} \tag{3.61}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \\ 1 - \frac{1}{2}b_h \geq& 1 \\ b_h \leq& 0 \end{aligned} \tag{3.62}$$

In neither treatment will t_1 choose b_h, as b_h must be positive.

$\mathbf{s_i} = (\mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1})$

Audience's belief

$$\begin{aligned} Pr(o_h, t_1) =& pp\frac{1}{2} + p(1-p) = p - \frac{1}{2}p^2 \\ Pr(o_h, t_2) =& (1-p)^2\frac{1}{2} \\ Pr(t_1|o_h) =& \frac{p - \frac{1}{2}p^2}{p - \frac{1}{2}p^2 + (1-p)^2\frac{1}{2}} = (2p - p^2) \\ E[t_1|o_h] =& (2p - p^2) \end{aligned} \tag{3.63}$$

Dictator Decision: t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) - \frac{2}{3}\sigma(1 - (2p - p^2)) = \frac{1}{3}\frac{1}{2} - \frac{2}{3}\sigma(p-1)^2 = \frac{1}{6} - \frac{2}{3}\sigma(p-1)^2 \tag{3.64}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma(1 - 0) = \frac{1}{3}(1 - o_l) - \frac{2}{3}\sigma = \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \tag{3.65}$$

$$
\begin{aligned}
U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) &> U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) \\
\frac{1}{6} - \frac{2}{3}\sigma(p-1)^2 &> \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \\
p > 1 - \frac{1}{2}\sqrt{3+2o_l} & \quad (public\ \sigma = 1) \\
p > 1 - \frac{\sqrt{-3+10o_l}}{\sqrt{2}} \quad for\ o_l \geq \frac{3}{10} & \quad (private\ \sigma = \frac{1}{10})
\end{aligned}
\tag{3.66}
$$

In the *public* treatment $p > 1-\frac{1}{2}\sqrt{3+2o_l}$. In the *private* treatment $o_l > \frac{3}{10}$ and $p > 1-\frac{\sqrt{-3+10o_l}}{\sqrt{2}}$.

Auction Decision: t_1 chooses b_h, if

$$
\begin{aligned}
U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\
&+ (1-p)\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\
=& 1 + \frac{1}{2}p(-b_h) \\
=& 1 - \frac{1}{2}b_h p
\end{aligned}
\tag{3.67}
$$

$$
\begin{aligned}
U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\
&+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_6^j), \sigma, t_1)\Big)\Big] \\
=& 1
\end{aligned}
\tag{3.68}
$$

$$
\begin{aligned}
U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \\
1 - \frac{1}{2}b_h p \geq& 1 \\
b_h p \leq& 0
\end{aligned}
\tag{3.69}
$$

In neither treatment t_1 will choose b_h as b_h and p are both positive.

$\mathbf{s_i = (0, 1, 1, 1, 1, 1)}$

Audience's belief

$$
\begin{aligned}
Pr(o_h, t_1) =& \frac{1}{2}p^2 \\
Pr(o_h, t_2) =& (1-p)p + \frac{1}{2}(1-p)^2 = \frac{1}{2}(1-p^2) \\
Pr(t_1|o_h) =& \frac{\frac{1}{2}p^2}{\frac{1}{2}p^2 + \frac{1}{2}(1-p^2)} = p^2 \\
E[t_1|o_h] =& p^2
\end{aligned}
\tag{3.70}
$$

Dictator Decision: t_2 chooses o_h, if

$$
U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) - \frac{2}{3}\sigma(1-p^2) = \frac{1}{3}\frac{1}{2} - \frac{2}{3}\sigma(1-p^2) = \frac{1}{6} - \frac{2}{3}\sigma(1-p^2) \tag{3.71}
$$

$$
U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma(1-0) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma = \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \tag{3.72}
$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) &> U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) \\ \frac{1}{6} - \frac{2}{3}\sigma(1-p^2) &> \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \\ p &> \frac{1}{2}\sqrt{1-2o_l} \quad (\textit{public } \sigma = 1) \\ p > \frac{\sqrt{5-10o_l}}{\sqrt{2}} \quad \textit{for } o_l &> \frac{3}{10} \quad (\textit{private } \sigma = \frac{1}{10}) \end{aligned} \tag{3.73}$$

In the *public* treatment $p > \frac{1}{2}\sqrt{1-2o_l}$. In the *private* treatment $o_l > \frac{3}{10}$ and $p > \frac{\sqrt{5-10o_l}}{\sqrt{2}}$.

Auction Decision: t_1 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big)\Big] \\ =& 1 + \frac{1}{2}(1-p)(-b_h) \\ =& 1 - \frac{1}{2}(1-p)b_h \end{aligned} \tag{3.74}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big] \\ &+ (1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_1)\Big)\Big] \\ =& 1 \end{aligned} \tag{3.75}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \\ 1 \geq& 1 - \frac{1}{2}(1-p)b_h \\ b_h \geq& b_h p \end{aligned} \tag{3.76}$$

In either treatment for $p \leq 1$ and $b_h > 0$ t_1 will choose b_l.

Auction Decision: t_2 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\ =& 1 + p(\frac{1}{6} - \frac{2}{3}\sigma(1-p^2)) + \frac{1}{2}(1-p)(\frac{1}{6} - \frac{2}{3}\sigma(1-p^2) - b_h + \frac{1}{6}) \\ =& \frac{1}{6}(7 + 3b_h(p-1) + 2(p-1)(1+p)^2\sigma) \end{aligned} \tag{3.77}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_4^j), \sigma, t_2)\Big)\Big] \\ &+ (1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\ =& 1 + \frac{1}{2}p(\frac{1}{6} - \frac{2}{3}\sigma(1-p^2) + \frac{1}{6}) + (1-p)(\frac{1}{6}) \\ =& \frac{1}{6}(7 - 2p\sigma + 2p^3\sigma) \end{aligned} \tag{3.78}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) &\geq U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \\ \frac{1}{6}(7 + 3b_h(p-1) + 2(p-1)(1+p)^2\sigma) &\geq \frac{1}{6}(7 - 2p\sigma + 2p^3\sigma) \\ \frac{1}{2}(-2 - 3b_h) \geq p \geq 1 \quad if\ b_h &> -\frac{4}{3} \quad (public\ \sigma = 1) \\ -1 - 15b_h \geq p \geq 1 \quad if\ b_h &> -\frac{2}{15} \quad (private\ \sigma = 0) \end{aligned} \tag{3.79}$$

In neither treatment t_2 will choose b_h, because $0 < p < 1$

$\mathbf{s_i = (0, 0, 1, 1, 1, 1)}$

Audience's belief

$$\begin{aligned} Pr(o_h, t_1) =& \frac{1}{2}p^2 + \frac{1}{2}p(1-p) = \frac{1}{2}p \\ Pr(o_h, t_2) =& \frac{1}{2}(1-p)p + \frac{1}{2}(1-p)^2 = \frac{1}{2}(1-p) \\ Pr(t_1|o_h) =& \frac{\frac{1}{2}p}{\frac{1}{2}p + \frac{1}{2}(1-p)} = p \\ E[t_1|o_h] =& p \end{aligned} \tag{3.80}$$

Dictator Decision: t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) - \frac{2}{3}\sigma(1-p) = \frac{1}{3}\frac{1}{2} - \frac{2}{3}\sigma(1-p) = \frac{1}{6} - \frac{2}{3}\sigma(1-p) \tag{3.81}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma(1-0) = \frac{1}{3}(S - o_l) - \frac{2}{3}\sigma = \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \tag{3.82}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) &> U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) \\ \frac{1}{6} - \frac{2}{3}\sigma(1-p) &> \frac{1}{3} - \frac{1}{3}o_l - \frac{2}{3}\sigma \\ p &> \frac{1}{4}(1 - 2o_l) \quad (public\ \sigma = 1) \\ p > \frac{1}{2}(5 - 10o_l) \quad for\ o_l &> \frac{3}{10} \quad (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.83}$$

In the *public* treatment $p > \frac{1}{4}(1 - 2o_l)$. In the *private* treatment $o_l > \frac{3}{10}$ and $p > \frac{1}{2}(5 - 10o_l)$.

Auction Decision t_1 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ =& 1 \end{aligned} \tag{3.84}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_6^j), \sigma, t_1)\Big)\Big] \\ =& 1 \end{aligned} \tag{3.85}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \geq & U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \\ 1 \geq & 1 \end{aligned} \tag{3.86}$$

In both treatments t_1 will choose b_l.

Auction Decision: t_2 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ =& 1 + p(\frac{1}{6} - \frac{2}{3}\sigma(1-p)) + (1-p)(\frac{1}{6} - \frac{2}{3}\sigma(1-p)) \\ =& \frac{1}{6}(7 + 4(-1+p)\sigma) \end{aligned} \tag{3.87}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_4^j), \sigma, t_2)\Big)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_6^j), \sigma, t_2))\Big)\Big] \\ =& 1 + \frac{1}{2}p(\frac{1}{6} - \frac{2}{3}\sigma(1-p) + \frac{1}{6}) + \frac{1}{2}(1-p)(\frac{1}{6} - \frac{2}{3}\sigma(1-p) + \frac{1}{6}) \\ =& \frac{1}{6}(7 + 2(-1+p)\sigma) \end{aligned} \tag{3.88}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \geq & U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \\ \frac{1}{6}(7 + 2(-1+p)\sigma) \geq & \frac{1}{6}(7 + 4(-1+p)\sigma) \\ p \leq & 1 \quad (public\ \sigma = 1) \\ p \leq & 1 \quad (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.89}$$

In both treatments t_2 will choose b_l.

A.3 Robustness Checks

Pure Selfishness

Purely selfish types $t_3 = (0, 0)$ have neither a fairness concern nor an image concern. They are only interested in maximizing their financial gains. Let the common prior about such a type be $Pr(t = t_3) = 1 - p$. The following expression gives their utility function in the Dictator game.

$$\begin{aligned} U_i^D(\alpha_i(I_\tau^i); \sigma, t_i(\theta_i,\mu_i)) =& \alpha_i(I_\tau^i)\Big[(1 - \theta_i - \mu_i)(S - o_h) - \theta_i(o^{eq} - o_h) - \mu_i\sigma(\bar{\theta} - E[\theta|o_h])\Big] \\ &+ (1 - \alpha_i(I_\tau^i))\Big[(1 - \theta_i - \mu_i)(S - o_l) - \theta_i(o^{eq} - o_l) - \mu_i\sigma(\bar{\theta} - E[\theta|o_l])\Big] \\ =& \alpha_i(I_\tau^i)\Big[(S - o_h)\Big] + (1 - \alpha_i(I_\tau^i))\Big[(S - o_l)\Big] \end{aligned} \tag{3.90}$$

Type $t_3 = (0, 0)$ is only interested in financial gains. When choosing between o_h and o_l, they always prefer o_l. For $\kappa = 5, 6$

$$U_1^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_3) = (S - o_h) = \frac{1}{2} \tag{3.91}$$

$$U_1^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_3) = (S - o_l) = 1 - o_l \tag{3.92}$$

$$\begin{aligned} U_1^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_3) \geq & U_1^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_3) \\ 1 - o_l \geq & \frac{1}{2} \\ o_l \leq & \frac{1}{2} \end{aligned} \tag{3.93}$$

By definition $o_l < \frac{1}{2}$, and therefore type t_3 always chooses o_l. Since type t_1 always chooses o_h, non of the pooling strategies qualify as an equilibrium. The four candidates left are all separating equilibria.
In each of these four profiles, the audience's belief is the same. Whenever they observe player i choosing o_h they know for sure that player i is a t_1 type $E[t_1|o_h] = 1$, because only a t_1 type chooses o_h.

$\mathbf{s_i = (1, 1, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_h, if The t_1 type does not care about image concerns, and therefore only their opponent's action enters their utility in the auction decision. It turns out that the same condition applies as when there exists a type t_2.

In either treatment for $b_h < \frac{1}{2} - o_l$ the fraction is positive. For $p < \frac{1-2b_h-2o_l}{1-2o_l}$, t_1 chooses b_h.

Auction Decision: t_3 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_3) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_3) - b_h) + \frac{1}{2}(U_1^R(\alpha_j(I_3^j), \sigma, t_3)\Big)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_1^D(\alpha_i(I_5^i); \sigma, t_3) - b_h) + \frac{1}{2}U_1^R(\alpha_j(I_5^j), \sigma, t_3)\Big)\Big] \\ =& 1 + \frac{1}{2}p\Big(1 - o_l - b_h + \frac{1}{2}\Big) + \frac{1}{2}(1-p)\Big(1 - o_l - b_h + o_l\Big) \\ =& \frac{1}{4}\Big(6 - 2b_h + p - 2po_l\Big) \end{aligned} \tag{3.94}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_3) =& 1 + p\Big[\Big(U_1^R(\alpha_j(I_3^j); \sigma, t_3)\Big)\Big] + (1-p))\Big[\Big(U_1^R(\alpha_j(I_5^j), \sigma, t_3)\Big)\Big] \\ =& 1 + p\Big(\frac{1}{2}\Big) + (1-p)\Big(o_l\Big) \\ =& 1 + p\Big(\frac{1}{2} - o_l\Big) + o_l \end{aligned} \tag{3.95}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_3) \geq & U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_3) \\ \frac{1}{4}\Big(6 - 2b_h + p - 2po_l\Big) \geq & 1 + p\Big(\frac{1}{2} - o_l\Big) + o_l \\ p \leq & \frac{2 - 2b_h - 4o_l}{1 - 2o_l} \end{aligned} \tag{3.96}$$

In either treatment for $b_h < 1 - 2o_l$ the fraction is greater than 0, t_3 chooses b_h. For $p \leq \frac{2-2b_h-4o_l}{1-2o_l}$.

$\mathbf{s_i = (1, 0, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_h, if
In either treatment for any $b_h > 0$ and $p \leq \frac{1-2o_l}{2+2b_h-4o_l}$, t_1 chooses b_h.

Auction Decision: t_3 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_3) =& 1 + p\Big[\Big(\frac{1}{2}(U_1^D(\alpha_i(I_5^i); \sigma, t_3) - b_h) + \frac{1}{2}(U_1^R(\alpha_j(I_3^j), \sigma, t_3)\Big)\Big] \\ &+ (1-p)\Big[\Big(U_1^D(\alpha_i(I_5^i); \sigma, t_3) - b_l\Big)\Big] \\ =& 1 + \frac{1}{2}p(1 - o_l - b_h + \frac{1}{2}) + (1-p)(1-o_l) \\ =& 2 - \frac{1}{4}p(1 + 2b_h - 2o_l) - o_l \end{aligned} \tag{3.97}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_3) =& 1 + p\Big[\Big(U_1^R(\alpha_j(I_3^j); \sigma, t_3)\Big) + \Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_1^D(\alpha_i(I_6^i); \sigma, t_3) - b_l) + \frac{1}{2}(U_1^R(\alpha_j(I_6^j), \sigma, t_3))\Big)\Big] \\ =& 1 + p(\frac{1}{2}) + \frac{1}{2}(1-p)(1 - o_l + o_l) \\ =& \frac{3}{2} \end{aligned} \tag{3.98}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_3) \geq& U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_3) \\ \frac{3}{2} \geq& 2 - \frac{1}{4}p(1 + 2b_h - 2o_l) - o_l \\ p \geq& \frac{2 - 4o_l}{1 + 2b_h - 2o_l} \end{aligned} \tag{3.99}$$

In the *private* treatment for $bh > \frac{1}{2}(1 - 2o_l)$ the fraction is positive and smaller 1. For $p \geq \frac{2-4o_l}{1+2b_h-2o_l}$, t_3 chooses b_l.

$\mathbf{s_i = (0, 1, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_l, if
In either treatment for $b_h > \frac{1}{2} - o_l$ and any p, t_1 chooses b_l.

Auction Decision: t_3 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_3) =& 1 + p\Big[\Big(U_1^D(\alpha_i(I_5^i); \sigma, t_3) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_1^D(\alpha_i(I_5^i); \sigma, t_3) - b_h) + \frac{1}{2}U_1^R(\alpha_j(I_5^j), \sigma, t_3)\Big)\Big] \\ =& 1 + p(1 - o_l) + \frac{1}{2}(1-p)(1 - o_l - b_h + o_l) \\ =& \frac{1}{2}(3 + b_h(-1+p) + p - 2po_l) \end{aligned} \tag{3.100}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_3) =& 1 + p\Big[\Big(\frac{1}{2}(U_1^D(\alpha_i(I_6^i); \sigma, t_3) - b_l) + \frac{1}{2}(U_1^R(\alpha_j(I_4^j), \sigma, t_3)\Big)\Big] \\ &+ (1-p)\Big[\Big(U_1^R(\alpha_j(I_5^j), \sigma, t_3)\Big)\Big] \\ =& 1 + \frac{1}{2}p(1 - o_l + \frac{1}{2}) + (1-p)(o_l) \\ =& 1 + o_l - \frac{3}{4}p(-1 + 2o_l) \end{aligned} \tag{3.101}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_3) \geq U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_3) \\ \frac{1}{2}(3 + b_h(-1+p) + p - 2po_l) \geq& 1 + o_l - \frac{3}{4}p(-1 + 2o_l) \\ p\begin{cases} < & \frac{2-2b_h-4o_l}{1-2b_h-2o_l} \quad if\ b_h < \frac{1}{2}(1-2o_l) \\ \geq & \frac{2-2b_h-4o_l}{1-2b_h-2o_l} \quad if\ b_h \geq \frac{1}{2}(1-2o_l) \end{cases} \end{aligned} \tag{3.102}$$

In either treatment for $b_h < \frac{1}{2}(1-2o_l)$ the fraction is positive and larger 1. For any p, t_2 chooses b_h. For $b_h \geq \frac{1}{2}(1-2o_l)$ and $p \geq \frac{2-2b_h-4o_l}{1-2b_h-2o_l}$, t_2 chooses b_h.

$\mathbf{s_i} = (\mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{1}, \mathbf{0}, \mathbf{0})$

Auction Decision: t_1 chooses b_l, if

In either treatment for $o_l < \frac{1}{2}$ p would need to be greater than 1, therefore t_1 will not choose b_h.

Equilibrium Summary

- Separating Equilibria
  - $s_i = (1, 1, 1, 1, 0, 0)$ in either treatment under conditions
    - For $b_h < \frac{1}{2} - o_l$ and $p < \frac{1-2b_h-2o_l}{1-2o_l}$
  - $s_i = (0, 1, 1, 1, 0, 0)$ in either treatment under conditions
    - For $b_h > \frac{1}{2} - o_l$ and $p \geq \frac{2-2bh-4xl}{1-2bh-2xl}$

Image Concerns as Benefits

Image concerns can also be modeled as benefits. This paradigm suggests that image concerns motivate individuals to shine in front of others by adhering to the norm, rather than feeling ashamed when they don't. An unfavorable image has no further effect on utility, it is a neutral signal. A favorable image instead increases utility. We assume a t_2 type under image concerns as benefits to exhibit the same preferences as before ($t_2 = (0, \frac{2}{3})$), but the dictator's utility function looks slightly different. The common prior on such a type is again $Pr(t = t_2) = 1 - p$.

$$\begin{aligned} U_i^D(\alpha_i(I_\tau^i);\sigma,t_i(\theta_i,\mu_i)) &= \alpha_i(I_\tau^i)\Big[(1-\theta_i-\mu_i)(S-o_h)-\theta_i(o^{eq}-o_h)+\mu_i\sigma E[\theta|o_h]\Big] \\ &+(1-\alpha_i(I_\tau^i))\Big[(1-\theta_i-\mu_i)(S-o_l)-\theta_i(o^{eq}-o_l)+\mu_i\sigma E[\theta|o_l]\Big] \\ &= \alpha_i(I_\tau^i)\Big[\frac{1}{3}(S-o_h)+\frac{2}{3}\sigma E[\theta|o_h]\Big] \\ &+(1-\alpha_i(I_\tau^i))\Big[\frac{1}{3}(S-o_l)+\frac{2}{3}\sigma E[\theta|o_l]\Big] \end{aligned} \tag{3.103}$$

Separating Equilibria

In any separating equilibrium, t_1 types choose o_h and t_2 types choose o_l. Therefore the audience knows that a signal o_h comes from a t_1-type and $E[t_1|o_h] = 1$ and $E[t_1|o_l] = 0$.

Dictator Decision t_2 chooses o_l, if
Let $\kappa = 5, 6$

$$U_1^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_3) = \frac{1}{3}(S-o_h) + \frac{2}{3}\sigma E[t_1|o_h] = \frac{1}{6} + \frac{2}{3}\sigma \tag{3.104}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S-o_l) + \frac{2}{3}\sigma E[t_1|o_l] = \frac{1}{3}(1-o_l) \tag{3.105}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 0;\sigma,t_2) &> U_i^D(\alpha_i(I_\kappa^i) = 1;\sigma,t_2) \\ \frac{1}{3}(1-o_l) &> \frac{1}{6}+\frac{2}{3}\sigma \\ o_l &< \frac{1}{2}-2\sigma \\ o_l < -\frac{3}{2} & \quad (public\ \sigma = 1) \\ o_l < \frac{3}{10} & \quad (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.106}$$

There is no separating equilibrium in the *public* treatment. For a separating equilibrium to exist, in the *private* treatment $o_l < \frac{3}{10}$.

$\mathbf{s_i = (1, 1, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_h, if
The t_1 type does not care about image concerns, and therefore only their opponent's action enters their utility in the auction decision. It turns out that in this scenario, the same conditions apply as when image concerns enter as benefits.
In the *private* treatment for $b_h < \frac{1}{2} - o_l$ the fraction is positive. For $p < \frac{1-2b_h-2o_l}{1-2o_l}$, t_1 chooses b_h.

Auction Decision: t_2 chooses b_h, if

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_2)\Big)\Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\ =& 1 + \frac{1}{2}p\Big(\frac{1}{3}(1-o_l) - b_h + \frac{1}{6}\Big) + \frac{1}{2}(1-p)\Big(\frac{1}{3}(1-o_l) - b_h + \frac{1}{3}o_l\Big) \\ =& \frac{1}{12}\Big(14 - 6b_h + p - 2po_l\Big)\end{aligned} \tag{3.107}$$

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_2)\Big)\Big] + (1-p))\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\ =& 1 + p\Big(\frac{1}{6}\Big) + (1-p)\Big(\frac{1}{3}o_l\Big) \\ =& \frac{1}{6}(6 + p + 2o_l - 2po_l)\end{aligned} \tag{3.108}$$

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \geq& U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \\ \frac{1}{12}\Big(14 - 6b_h + p - 2po_l\Big) \geq& \frac{1}{6}(6 + p + 2o_l - 2po_l) \\ p \leq& \frac{2 - 6b_h - 4o_l}{1 - 2o_l}\end{aligned} \tag{3.109}$$

In the *private* treatment for $bh < \frac{1}{3}(1 - 2o_l)$ the fraction is greater than 0, t_2 chooses b_h for $p \leq \frac{2-6b_h-4o_l}{1-2o_l}$.

$\mathbf{s_i = (1, 0, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_h, if
In the *private* treatment for any $b_h > 0$ and $p \leq \frac{1-2o_l}{2+2b_h-4o_l}$, t_1 chooses b_h.

Auction Decision: t_2 chooses b_l, if

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_2)\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ =& 1 + \frac{1}{2}p(\frac{1}{3}(1-o_l) - b_h + \frac{1}{6}) + (1-p)(\frac{1}{3}(1-o_l)) \\ =& \frac{1}{12}(-4(-4+o_l) + p(-1 - 6b_h + 2o_l))\end{aligned} \tag{3.110}$$

$$\begin{aligned}U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_2)\Big) + \Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_6^j), \sigma, t_2))\Big)\Big] \\ =& 1 + p(\frac{1}{6}) + \frac{1}{2}(1-p)(\frac{1}{3}(1-o_l) + \frac{1}{3}o_l) \\ =& \frac{7}{6}\end{aligned} \tag{3.111}$$

$$
\begin{aligned}
U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \geq & U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \\
\frac{7}{6} \geq & \frac{1}{12}(-4(-4 + o_l) + p(-1 - 6b_h + 2o_l)) \\
p \geq & \frac{2 - 4o_l}{1 + 6b_h - 2o_l}
\end{aligned} \tag{3.112}
$$

In the *private* treatment for $bh > \frac{1}{6}(1 - 2xl)$ the fraction is positive and smaller 1. For $p \geq \frac{2-4o_l}{1+6b_h-2o_l}$, t_2 chooses b_l.

$\mathbf{s_i = (0, 1, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_l, if

In the *private* treatment for $b_h > \frac{1}{2} - o_l$ and any p, t_1 chooses b_l.

Auction Decision: t_2 chooses b_h, if

$$
\begin{aligned}
U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) = & 1 + p\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\
& +(1 - p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\
= & 1 + p(\frac{1}{3}(1 - o_l)) + \frac{1}{2}(1 - p)(\frac{1}{3}(1 - o_l) - b_h + \frac{1}{3}o_l) \\
= & \frac{1}{6}(7 + 3b_h(-1 + p) + p - 2po_l)
\end{aligned} \tag{3.113}
$$

$$
\begin{aligned}
U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) = & 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_4^j), \sigma, t_2)\Big)\Big] \\
& +(1 - p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\
= & 1 + \frac{1}{2}p(\frac{1}{3}(1 - o_l) + \frac{1}{6}) + (1 - p)(\frac{1}{3}o_l) \\
= & \frac{1}{12}(p(3 - 6o_l) + 4(3 + o_l))
\end{aligned} \tag{3.114}
$$

$$
\begin{aligned}
U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \geq & \; U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \\
\frac{1}{6}(7 + 3b_h(-1 + p) + p - 2po_l) \geq & \frac{1}{12}(p(3 - 6o_l) + 4(3 + o_l)) \\
p \begin{cases} < & \frac{2-6b_h-4o_l}{1-6b_h-2o_l} \quad if\ b_h < \frac{1}{6}(1 - 2o_l) \\ \geq & \frac{2-6b_h-4o_l}{1-6b_h-2o_l} \quad if\ b_h \geq \frac{1}{6}(1 - 2o_l) \end{cases} &
\end{aligned} \tag{3.115}
$$

In the *private* treatment for $b_h < \frac{1}{6}(1 - 2o_l)$ the fraction is positive and larger 1. For any p, t_2 chooses b_h. For $b_h \geq \frac{1}{6}(1 - 2o_l)$ and $p \geq \frac{2-6b_h-4o_l}{1-6b_h-2o_l}$, t_2 chooses b_h.

$\mathbf{s_i = (0, 0, 1, 1, 0, 0)}$

Auction Decision: t_1 chooses b_l, if

In either treatment for $o_l < \frac{1}{2}$ p would need to be greater than 1, therefore t_1 will not choose b_h.

Pooling Equilibria

In any pooling equilibrium, t_1 types choose o_h and t_2 types choose o_h as well. Therefore, the audience does not know exactly the type after observing o_h.

$\mathbf{s_i} = (\mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1}, \mathbf{1})$

Audience's belief

$$\begin{aligned} Pr(o_h, t_1) =& pp\frac{1}{2} + p(1-p)\frac{1}{2} = p\frac{1}{2} \\ Pr(o_h, t_2) =& (1-p)p\frac{1}{2} + (1-p)(1-p)\frac{1}{2} = (1-p)\frac{1}{2} \\ Pr(t_1|o_h) =& \frac{p\frac{1}{2}}{p\frac{1}{2} + (1-p)\frac{1}{2}} = p \\ E[t_1|o_h] =& p \end{aligned} \tag{3.116}$$

Dictator Decision: t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h)\frac{2}{3}\sigma p = \frac{1}{3}\frac{1}{2} + \frac{2}{3}\sigma p = \frac{1}{6} + \frac{2}{3}\sigma p \tag{3.117}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) + \frac{2}{3}\sigma 0 = \frac{1}{3}(S - o_l) = \frac{1}{3} - \frac{1}{3}o_l \tag{3.118}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) \geq& U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) \\ \frac{1}{6} + \frac{2}{3}\sigma p \geq& \frac{1}{3} - \frac{1}{3}o_l \\ p \geq \frac{1}{4}(1 - 2o_l) \quad & (public\ \sigma = 1) \\ p \geq \frac{1}{2}(5 - 10o_l) \quad & (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.119}$$

In the *public* treatment for $p > \frac{1}{4}(1-2o_l)$. In the *private* treatment for $o_l > \frac{3}{10}$ and $p > \frac{1}{2}(5-10o_l)$.

Auction Decision: t_1 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\ &+(1-p))\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big)\Big] \\ =& 1 + \frac{1}{2}p(-b_h) + \frac{1}{2}(1-p)(-b_h) \\ =& 1 - \frac{1}{2}b_h \end{aligned} \tag{3.120}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^R(\alpha_j(I_3^j), \sigma, t_1)\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_1)\Big)\Big] \\ =& 1 \end{aligned} \tag{3.121}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i)=1,s_j;\sigma,t_1) \geq& U_i^A(\beta_i(I_1^i)=0,s_j;\sigma,t_1) \\ 1-\frac{1}{2}b_h \geq& 1 \\ b_h \leq& 0 \end{aligned} \tag{3.122}$$

In neither treatment t_1 will choose b_h as b_h must be positive.

$\mathbf{s_i = (1, 0, 1, 1, 1, 1)}$

Audience's belief

$$\begin{aligned} Pr(o_h,t_1) =& pp\frac{1}{2}+p(1-p) = p-\frac{1}{2}p^2 \\ Pr(o_h,t_2) =& (1-p)^2\frac{1}{2} \\ Pr(t_1|o_h) =& \frac{p-\frac{1}{2}p^2}{p-\frac{1}{2}p^2+(1-p)^2\frac{1}{2}} = (2p-p^2) \\ E[t_1|o_h] =& (2p-p^2) \end{aligned} \tag{3.123}$$

Dictator Decision: t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i)=1;\sigma,t_2) = \frac{1}{3}(S-o_h)+\frac{2}{3}\sigma(2p-p^2) = \frac{1}{3}\frac{1}{2}+\frac{2}{3}\sigma(2p-p^2) = \frac{1}{6}(1+8p\sigma-4p^2\sigma) \tag{3.124}$$

$$U_i^D(\alpha_i(I_\kappa^i)=0;\sigma,t_2) = \frac{1}{3}(S-o_l)+\frac{2}{3}\sigma 0 = \frac{1}{3}(1-o_l) = \frac{1}{3}-\frac{1}{3}o_l \tag{3.125}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i)=1;\sigma,t_2) \geq& U_i^D(\alpha_i(I_\kappa^i)=0;\sigma,t_2) \\ \frac{1}{6}(1+8p\sigma-4p^2\sigma) \geq& \frac{1}{3}-\frac{1}{3}o_l \\ p \geq 1-\frac{1}{2}\sqrt{3+2o_l} \quad& (public\ \sigma=1) \\ p \geq 1-\frac{\sqrt{-3+10o_l}}{\sqrt{2}} \quad for\ o_l > \frac{3}{10} \quad& (private\ \sigma=\frac{1}{10}) \end{aligned} \tag{3.126}$$

In the *public* treatment for $p > 1-\frac{1}{2}\sqrt{3+2o_l}$. In the *private* treatment for $o_l > \frac{3}{10}$ and $p > 1-\frac{\sqrt{-3+10o_l}}{\sqrt{2}}$.

Auction Decision: t_1 chooses b_h, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i)=1,s_j;\sigma,t_1) =& 1+p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i);\sigma,t_1)-b_h)+\frac{1}{2}(U_i^R(\alpha_j(I_3^j),\sigma,t_1)\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^D(\alpha_i(I_3^i);\sigma,t_1)-b_l\Big)\Big] \\ =& 1+\frac{1}{2}p(-b_h) \\ =& 1-\frac{1}{2}b_h p \end{aligned} \tag{3.127}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i)=0,s_j;\sigma,t_1) =& 1+p\Big[\Big(U_i^R(\alpha_j(I_3^j),\sigma,t_1)\Big)\Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i);\sigma,t_1)-b_l)+\frac{1}{2}U_i^R(\alpha_j(I_6^j),\sigma,t_1)\Big)\Big] \\ =& 1 \end{aligned} \tag{3.128}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i)=1,s_j;\sigma,t_1) \geq& U_i^A(\beta_i(I_1^i)=0,s_j;\sigma,t_1) \\ 1-\frac{1}{2}b_hp \geq& 1 \\ b_hp \leq& 0 \end{aligned} \tag{3.129}$$

In neither treatment will the t_1-type choose b_h, since both b_h and p are positive.

$\mathbf{s_i = (0,1,1,1,1,1)}$

Audience's belief

$$\begin{aligned} Pr(o_h,t_1) =& \frac{1}{2}p^2 \\ Pr(o_h,t_2) =& (1-p)p+\frac{1}{2}(1-p)^2=\frac{1}{2}(1-p^2) \\ Pr(t_1|o_h) =& \frac{\frac{1}{2}p^2}{\frac{1}{2}p^2+\frac{1}{2}(1-p^2)}=p^2 \\ E[t_1|o_h] =& p^2 \end{aligned} \tag{3.130}$$

Dictator Decision: t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i)=1;\sigma,t_2)=\frac{1}{3}(S-o_h)+\frac{2}{3}\sigma p^2=\frac{1}{3}\frac{1}{2}+\frac{2}{3}\sigma p^2=\frac{1}{6}+\frac{2}{3}\sigma p^2 \tag{3.131}$$

$$U_i^D(\alpha_i(I_\kappa^i)=0;\sigma,t_2)=\frac{1}{3}(S-o_l)\frac{2}{3}\sigma 0=\frac{1}{3}(S-o_l)=\frac{1}{3}-\frac{1}{3}o_l \tag{3.132}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i)=1;\sigma,t_2) \geq& U_i^D(\alpha_i(I_\kappa^i)=0;\sigma,t_2) \\ \frac{1}{6}+\frac{2}{3}\sigma p^2 \geq& \frac{1}{3}-\frac{1}{3}o_l \\ p \geq \frac{1}{2}\sqrt{1-2o_l} & \quad (public\ \sigma=1) \\ p \geq \frac{\sqrt{5-10o_l}}{\sqrt{2}} \quad for\ o_l>\frac{3}{10} & \quad (private\ \sigma=\frac{1}{10}) \end{aligned} \tag{3.133}$$

In the *public* treatment for $p>\frac{1}{2}\sqrt{1-2o_l}$. In the *private* treatment for $o_l>\frac{3}{10}$ and $p>\frac{\sqrt{5-10o_l}}{\sqrt{2}}$.

Auction Decision: t_1 chooses b_l, if

$$\begin{aligned}
U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\
&+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_h) + \frac{1}{2}(U_i^R(\alpha_j(I_5^j), \sigma, t_1))\Big)\Big] \\
=& 1 + \frac{1}{2}(1-p)(-b_h) \\
=& 1 - \frac{1}{2}(1-p)b_h
\end{aligned} \tag{3.134}$$

$$\begin{aligned}
U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big] \\
&+ (1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_1)\Big)\Big] \\
=& 1
\end{aligned} \tag{3.135}$$

$$\begin{aligned}
U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \\
1 \geq& 1 - \frac{1}{2}(1-p)b_h \\
b_h \geq& b_h p
\end{aligned} \tag{3.136}$$

In either treatment for $p \leq 1$ and $b_h > 0$ the t_1-type will choose b_l.

Auction Decision: t_2 chooses b_h, if

$$\begin{aligned}
U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\
&+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_h) + \frac{1}{2}U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\
=& 1 + p(\frac{1}{6} + \frac{2}{3}\sigma p^2) + \frac{1}{2}(1-p)(\frac{1}{6} + \frac{2}{3}\sigma p^2 - b_h + \frac{1}{6}) \\
=& \frac{1}{6}(7 + 3b_h(p-1) + 2p^2\sigma + 2p^3\sigma)
\end{aligned} \tag{3.137}$$

$$\begin{aligned}
U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_4^j), \sigma, t_2)\Big)\Big] \\
&+ (1-p)\Big[\Big(U_i^R(\alpha_j(I_5^j), \sigma, t_2)\Big)\Big] \\
=& 1 + \frac{1}{2}p(\frac{1}{6} + \frac{2}{3}\sigma p^2 + \frac{1}{6}) + (1-p)(\frac{1}{6}) \\
=& \frac{1}{6}(7 + 2p^3\sigma)
\end{aligned} \tag{3.138}$$

$$\begin{aligned}
U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \geq& \; U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) \\
\frac{1}{6}(7 + 3b_h(p-1) + 2p^2\sigma + 2p^3\sigma) \geq& \frac{1}{6}(7 + 2p^3\sigma) \\
p \geq& -\frac{3b_h}{4} + \frac{1}{4}\sqrt{3}\sqrt{8b_h + 3b_h^2} \quad (public\ \sigma = 1) \\
p \geq& -\frac{15b_h}{2} + \frac{1}{2}\sqrt{15}\sqrt{4b_h + 15b_h^2} \quad (private\ \sigma = \frac{1}{10})
\end{aligned} \tag{3.139}$$

In the *private* treatment for $p \geq -\frac{15b_h}{2} + \frac{1}{2}\sqrt{15}\sqrt{4b_h + 15b_h^2}$ the t_2-type would choose b_h. In the *public* treatment for $p \geq -\frac{3b_h}{4} + \frac{1}{4}\sqrt{3}\sqrt{8b_h + 3b_h^2}$ the t_2-type chooses b_h.

$\mathbf{s_i = (0, 0, 1, 1, 1, 1)}$

Audience's belief

$$\begin{aligned} Pr(o_h, t_1) =& \frac{1}{2}p^2 + \frac{1}{2}p(1-p) = \frac{1}{2}p \\ Pr(o_h, t_2) =& \frac{1}{2}(1-p)p + \frac{1}{2}(1-p)^2 = \frac{1}{2}(1-p) \\ Pr(t_1|o_h) =& \frac{\frac{1}{2}p}{\frac{1}{2}p + \frac{1}{2}(1-p)} = p \\ E[t_1|o_h] =& p \end{aligned} \tag{3.140}$$

Dictator Decision t_2 chooses o_h, if

$$U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) = \frac{1}{3}(S - o_h) + \frac{2}{3}\sigma p = \frac{1}{3}\frac{1}{2} + \frac{2}{3}\sigma p = \frac{1}{6} + \frac{2}{3}\sigma p \tag{3.141}$$

$$U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) = \frac{1}{3}(S - o_l) + \frac{2}{3}\sigma 0 = \frac{1}{3}(S - o_l) = \frac{1}{3} - \frac{1}{3}o_l \tag{3.142}$$

$$\begin{aligned} U_i^D(\alpha_i(I_\kappa^i) = 1; \sigma, t_2) > U_i^D(\alpha_i(I_\kappa^i) = 0; \sigma, t_2) \\ \frac{1}{6} + \frac{2}{3}\sigma p > \frac{1}{3} - \frac{1}{3}o_l \\ p > \frac{1}{4}(1 - 2o_l) \quad (public\ \sigma = 1) \\ p > \frac{1}{2}(5 - 10o_l) \quad for\ o_l > \frac{3}{10} \quad (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.143}$$

In the *public* treatment for $p > \frac{1}{4}(1 - 2o_l)$ the t_2-type chooses o_h. In the *private* treatment for $o_l > \frac{3}{10}$ and $p > \frac{1}{2}(5 - 10o_l)$ the t_2-type chooses o_h.

Auction Decision t_1 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ &+ (1-p)\Big[\Big(U_i^D(\alpha_i(I_3^i); \sigma, t_1) - b_l\Big)\Big] \\ =& 1 \end{aligned} \tag{3.144}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_4^j), \sigma, t_1)\Big] \\ &+ (1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_4^i); \sigma, t_1) - b_l) + \frac{1}{2}U_i^R(\alpha_j(I_6^j), \sigma, t_1)\Big)\Big] \\ =& 1 \end{aligned} \tag{3.145}$$

$$\begin{aligned} U_i^A(\beta_i(I_1^i) = 1, s_j; \sigma, t_1) \geq& U_i^A(\beta_i(I_1^i) = 0, s_j; \sigma, t_1) \\ 1 \geq& 1 \end{aligned} \tag{3.146}$$

In both treatments, the t_1-type will choose b_l.

Auction Decision t_2 chooses b_l, if

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) =& 1 + p\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ &+(1-p)\Big[\Big(U_i^D(\alpha_i(I_5^i); \sigma, t_2) - b_l\Big)\Big] \\ =& 1 + p(\frac{1}{6} + \frac{2}{3}\sigma p) + (1-p)(\frac{1}{6} + \frac{2}{3}\sigma p) \\ =& \frac{1}{6}(7 + 4p\sigma) \end{aligned} \tag{3.147}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 0, s_j; \sigma, t_2) =& 1 + p\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_4^j), \sigma, t_2)\Big)\Big] \\ &+(1-p)\Big[\Big(\frac{1}{2}(U_i^D(\alpha_i(I_6^i); \sigma, t_2) - b_l) + \frac{1}{2}(U_i^R(\alpha_j(I_6^j), \sigma, t_2))\Big)\Big] \\ =& 1 + \frac{1}{2}p(\frac{1}{6} + \frac{2}{3}\sigma p + \frac{1}{6}) + \frac{1}{2}(1-p)(\frac{1}{6} + \frac{2}{3}\sigma p + \frac{1}{6}) \\ =& \frac{1}{6}(7 + 2p\sigma) \end{aligned} \tag{3.148}$$

$$\begin{aligned} U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \geq& U_i^A(\beta_i(I_2^i) = 1, s_j; \sigma, t_2) \\ \frac{1}{6}(7 + 2p\sigma)\sigma) \geq& \frac{1}{6}(7 + 4p\sigma) \\ p \leq& 0 \quad (public\ \sigma = 1) \\ p \leq& 0 \quad (private\ \sigma = \frac{1}{10}) \end{aligned} \tag{3.149}$$

In neither treatment the t_2-type chooses b_l.

Equilibrium Summary

- Separating Equilibria
  - $s_i = (1, 1, 1, 1, 0, 0)$ in the *private* treatment under conditions
    - $o_l < \frac{3}{10}$
    - For $b_h < \frac{1}{4}(1 - 2o_l)$ and $p < \frac{1-2b_h-2o_l}{1-2o_l}$
    - For $\frac{1}{4}(1 - 2o_l) < b_h < \frac{1}{3}(1 - 2o_l)$ and $p \leq \frac{2-6b_h-4o_l}{1-2o_l}$
  - $s_i = (1, 0, 1, 1, 0, 0)$ in the *private* treatment under conditions
    - $o_l < \frac{3}{10}$
    - for $b_h > \frac{1}{2}(3 - 6o_l)$ and $\frac{2-4o_l}{1+6b_h-2o_l} \leq p \leq \frac{1-2o_l}{2+2b_h-4o_l}$
  - $s_i = (0, 1, 1, 1, 0, 0)$ in the *private* treatment under conditions
    - $o_l < \frac{3}{10}$
    - For $b_h > \frac{1}{2} - o_l$ and $p \geq \frac{2-6b_h-4o_l}{1-6b_h-2o_l}$
- Pooling Equilibria
  - $s_i = (0, 1, 1, 1, 0, 0)$ in the *public* treatment under conditions
    - for $b_h \leq -\frac{4o_l-2}{9+6o_l} + \frac{1}{3}\sqrt{\frac{1-6o_l+12o_l^2-8o_l^3}{(3+2o_l)^2}}$ and $p \geq \frac{1}{2}\sqrt{1 - 2o_l}$
    - for $b_h > -\frac{4o_l-2}{9+6o_l} + \frac{1}{3}\sqrt{\frac{1-6o_l+12o_l^2-8o_l^3}{(3+2o_l)^2}}$ and $p \geq -\frac{3b_h}{4} + \frac{1}{4}\sqrt{3}\sqrt{8b_h + 3b_h^2}$
  - $s_i = (0, 1, 1, 1, 0, 0)$ in the *private* treatment under conditions
    - $o_l > \frac{3}{10}$

* for $o_l \leq \frac{1}{10}(5 - 30b_h - 117b_h^2) + \frac{3}{2}\sqrt{\frac{3}{5}}\sqrt{4b_h^3 + 15b_h^4}$ and $p > \frac{\sqrt{5-10o_l}}{\sqrt{2}}$
* for $o_l > \frac{1}{10}(5-30b_h-117b_h^2)+\frac{3}{2}\sqrt{\frac{3}{5}}\sqrt{4b_h^3 + 15b_h^4}$ and $p \geq -\frac{15b_h}{2}+\frac{1}{2}\sqrt{15}\sqrt{4b_h + 15b_h^2}$
  the t_2-type would choose b_h

B Additional Tables

Table B1: Auction Bids

	Tobit Estimates					
	Dependent Variable: Auction Bid					
	(1)	(2)	(3)	(4)	(5)	(6)
Public	-10.03	-22.42*	-29.11**			
	(7.044)	(12.64)	(13.88)			
Fairness	-0.962**	-1.447***	-0.888**	-1.115*	-1.909**	-1.148**
	(0.380)	(0.462)	(0.373)	(0.589)	(0.727)	(0.575)
Exp Offer	-0.448**	-0.424*	-0.784***	0.0120	0.0249	-0.126
	(0.214)	(0.216)	(0.297)	(0.241)	(0.243)	(0.275)
Public × Fairness		0.960				
		(0.735)				
Public × Exp Offer			0.640			
			(0.398)			
Image				-36.36***	-60.09***	-66.54***
				(10.65)	(14.80)	(22.57)
Image × Fairness					2.372**	
					(1.129)	
Image × Exp Offer						0.748
						(0.479)
Constant	79.54***	85.53***	86.41***	65.41***	75.30***	70.76***
	(7.752)	(8.602)	(8.755)	(11.92)	(12.45)	(12.49)
N	162	162	162	82	82	82
Pseudo R^2	0.0149	0.0161	0.0171	0.0203	0.0271	0.0223

Note: *Public* is a dummy variable for the treatment; *fairness* is the offer made in Part 1 (0 to 40); *exp offer* is the first-order belief about others' offers; *image* is a dummy variable for self-reported treatment influence; Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Tobit model: lower boundary - 0 and upper boundary - 100.

Table B2: Auction Bids - Photo Treatments

	OLS Estimates Dependent Variable: Auction Bid					
	(1)	(2)	(3)	(4)	(5)	(6)
Public	-17.94***	-24.50***	-7.771			
	(5.564)	(9.345)	(10.87)			
Fairness	-1.339***	-1.614***	-1.323***	-1.026**	-1.952***	-0.892*
	(0.330)	(0.416)	(0.329)	(0.470)	(0.657)	(0.513)
Exp Offer	-0.474**	-0.476**	-0.295	-0.682**	-0.564**	-0.985***
	(0.182)	(0.183)	(0.222)	(0.264)	(0.259)	(0.361)
Public × Fairness		0.492				
		(0.611)				
Public × Fairness			-0.357			
			(0.340)			
Image				-10.61	-37.41***	-39.36**
				(7.747)	(12.20)	(15.50)
Image × Fairness					2.113**	
					(0.852)	
Image × Exp Offer						0.873*
						(0.514)
Constant	90.09***	93.97***	85.53***	79.78***	87.75***	87.81***
	(5.976)	(6.931)	(6.925)	(9.219)	(9.571)	(9.455)
N	144	144	144	78	78	78
R^2	0.244	0.247	0.250	0.205	0.260	0.236

Note: *Public* is a dummy variable for the treatment; *Fairness* is the offer made in Part 1 (0 to 40); *Exp Offer* is the first-order belief about others' offers; *Image* is a dummy variable for self-reported treatment influence; Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3: Auction Bids - Photo Treatments

	Tobit Estimates					
	Dependent Variable: Auction Bid					
	(1)	(2)	(3)	(4)	(5)	(6)
Public	-24.59***	-31.31**	-4.592			
	(7.616)	(13.26)	(14.45)			
Fairness	-1.902***	-2.184***	-1.871***	-1.579**	-2.774***	-1.400*
	(0.458)	(0.625)	(0.455)	(0.669)	(0.947)	(0.715)
Exp Offer	-0.632**	-0.633**	-0.284	-1.090***	-0.927**	-1.488***
	(0.246)	(0.246)	(0.290)	(0.394)	(0.383)	(0.549)
Public × Fairness		0.502				
		(0.866)				
Public × Exp Offer			-0.705			
			(0.468)			
Image				-10.33	-44.22***	-47.47**
				(11.39)	(15.90)	(22.09)
Image × Fairness					2.735**	
					(1.210)	
Image × Exp Offer						1.145
						(0.757)
Constant	104.1***	108.2***	95.29***	94.37***	104.3***	105.0***
	(8.638)	(10.86)	(9.293)	(13.69)	(14.20)	(15.02)
N	144	144	144	78	78	78
Pseudo R^2	0.0356	0.0359	0.0378	0.0318	0.0390	0.0358

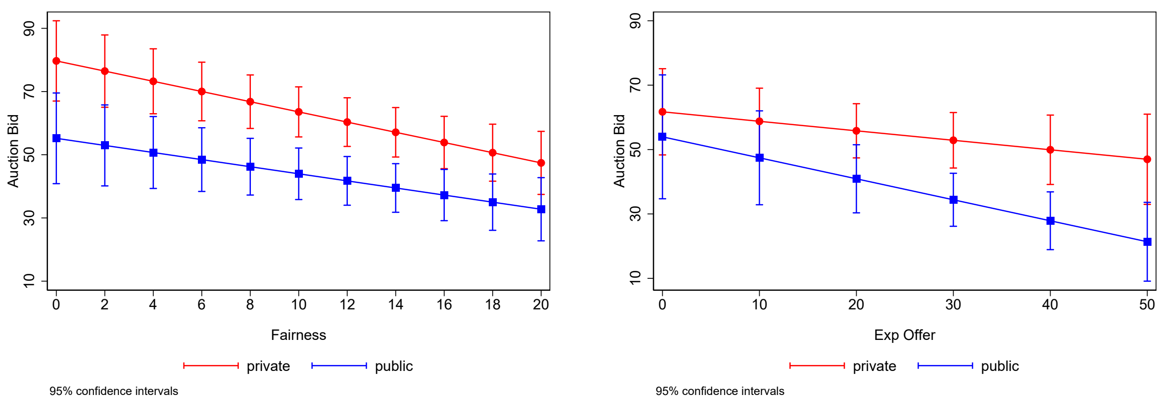
Note: *Public* is a dummy variable for the treatment; *Fairness* is the offer made in Part 1 (0 to 40); *Exp Offer* is the first-order belief about others' offers; *Image* is a dummy variable for self-reported treatment influence; Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Tobit model: lower boundary 0 and upper boundary 100.

Table B4: Auction Bids - Photo Treatments - New Image Index

	OLS & Tobit Estimates			
	Dependent Variable: Auction Bid			
	OLS	OLS	Tobit	Tobit
	(1)	(2)	(3)	(4)
Public	-18.03***	-16.96***	-24.70***	-22.81***
	(5.570)	(6.342)	(7.616)	(8.456)
Fairness	-1.335***	-1.334***	-1.897***	-1.896***
	(0.330)	(0.330)	(0.457)	(0.454)
Exp Offer	-0.470**	-0.467**	-0.628**	-0.623**
	(0.181)	(0.184)	(0.245)	(0.247)
High Image	-3.146	-1.062	-3.346	0.291
	(6.108)	(8.528)	(8.451)	(10.60)
Public × High Image		-3.933		-7.092
		(12.40)		(17.01)
Constant	90.83***	90.15***	104.9***	103.7***
	(6.228)	(6.775)	(8.914)	(9.356)
N	144	144	144	144
R^2	0.245	0.246	0.0358	0.0359

Note: Standard errors in parentheses; *Exp Offer* is the first-order belief about others' offers; *High Image* is a dummy variable for the image index to take on values above a certain threshold. Robust standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B1: Predicted Auction Bids - Photo Treatments



C Experimental Instructions

Welcome! Today you will participate in an economic experiment. The duration of the experiment will be around 45 minutes. During this time we kindly ask you not to communicate with each other in any way. Questions should be directed only at members of the staff. Please raise your hand and we will be happy to come to your cubicle to answer any questions. Please also keep your cubicle's curtains closed. It is mandatory that you switch off your phones for the course of the experiment. Should you fail to comply with the laboratory's guidelines this will lead to you being excluded from the experiment without pay. At the end of the experiment, we kindly ask you to answer a short socio-economic questionnaire. The analysis of the data from this experiment will be completely anonymized such that none of your decisions can be linked to your personal data afterwards.

Please note that today there will be a slight deviation from the usual payment procedure. A part of your earnings from the experiment may be made visible to other participants of this session. It will be made clear which part of the experiment is concerned. If there are no further questions or concerns regarding the procedure we can now start with the experiment.

By participating in the experiment you will earn real money. How much you earn depends on your decisions and the decisions of other participants. Therefore, please read and follow the instructions carefully. Over the course of the experiment, you will earn or lose points. At the end of the session, your points will be converted into money by a ratio of 18 points for 1 €. The following experiment has **three parts**. The instructions for Part 2 and 3 will be handed out after Part 1. Please note that your choices in Part 1 does not influence the procedure and your potential earnings in the following parts in any way.

On several occasions today you will be asked to **make a guess about what the other participants are doing**. When asked to do so, please enter your personal assessment of the average behavior of the other participants in this session. At the end, one of those questions will be randomly picked and your answer is compared to the true observed behavior. Your earnings from this question will be determined in the following way:

You earn 18 points if your answer is in an interval of +/- 10 percent around the actual average. You earn 9 points if your answer is in an interval of +/- 20 percent around the actual average. You do not earn any points for this question if your answer is further away from the actual average. To summarize: The closer your guess is to the actual average, the more points will you earn for this question type.

Part 1

In Part 1, you will be randomly matched in pairs without knowing who your partner is. In these pairs you will play a game that has to roles: Player S (sender/she) and player R (recipient/he).

The game is as follows: **Player S receives 40 points and can distribute them freely between herself and Player R.** Player R only receives as many points as Player S is sending to him. Player S keeps the rest of the points for herself.

Your points from Part 1 are thus calculated as follows:
Points for Player S = 40 - points sent to Player E
Points for Player E = 0 + points sent from Player S

You will now have to decide how many points you want to send in case you become Player S.

After every participant made this choice, the computer will randomly put together player pairs and again randomly decide who is Player S and who is Player E in each pair. Both participants in a pair are equally likely to become Player S or E. Your previous decision does not influence the matching of pairs or the role allocation. Player S's decision will be implemented in every pair.

You will learn at the end of today's session whether you were Player S or E and how many points you earned from Part 1. Please note that your earnings from Part 1 will never be made public.

Part 2 and 3

For Part 2 and 3 you will again be randomly put in pairs with another participant without knowing who he or she is. It is highly unlikely that you will be in a pair with the same person as in Part 1 and it is not possible to identify your partner.

In Part 3 you will again play a game with your partner which has to roles: Player A and Player B. In contrast to Part 1, it will be determined first who takes on which role. This is done in Part 2.

Part 2

Every player is now a part of a pair and receives 100 points. You can freely choose how many of these points (between 0 and 100) you want to use to increase your chances of becoming Player A in Part 3. Your partner simultaneously chooses an amount between 0 and 100 points. **The player who chose the higher amount of points will pay the amount of points which the other player has chosen.** The amount you pay is therefore never higher than the amount you chose. The player who chose the lower amount pays nothing and keeps the entire 100 points.

Your points from Part 2 are calculated as follows:
Player 1 chooses a higher amount of points than Player 2:
Points for Player 1 = 100 - chosen amount of Player 2
Points for Player 2 = 100

You will be informed about who chose more points in your pair at the end of the session. In the unlikely case where both of you chose exactly the same amount, the computer will break the tie and determine with equal chances who pays this amount.

Afterwards, every pair will take part in a lottery. The lottery is equivalent to blindly drawing a ball from an urn. There are nine blue balls and one red ball in the urn. If a blue ball is drawn, the player who chose the higher amount of points becomes Player A. If the red ball is drawn, the player who chose the lower amount becomes Player A. In both cases, the other player becomes Player B. The player who chose the higher amount therefore becomes Player A in 90 percent of the cases and the player who chose the lower amount in 10 percent of the cases. **Both players may become Player A, but it is more likely for the player who chose the higher amount of points.**

After the lottery every pair has a Player A and a Player B. The computer carries out the lottery in secret and will inform about the result, but not whether you or your partner chose more points.

Example 1:

- Player 1 chooses 47 points and Player 2 chooses 23 points.
- Player 1 chose the higher amount and pays 23 points (Player 2's amount). Player 2 pays nothing.

⇒ Player 1 keeps 77 points from Part 2 (100 - 23) and Player 2 keeps 100 points.

- In the lottery, the role of Player A will go to
  - Player 1 if a blue ball is drawn (9 out of 10 times)
  - Player 2 if the red ball is drawn (1 out of 10 times)

Beispiel 2:

- Player 1 chooses 0 points and Player 2 chooses 89 points.
- Player 2 chose the higher amount and pays 0 points (Player 1's amount). Player 2 pays nothing as well.

⇒ Both players keep 100 points from Part 2.

- In the lottery, the role of Player A will go to
  - Player 2 if a blue ball is drawn (9 out of 10 times)
  - Player 1 if the red ball is drawn (1 out of 10 times)

Part 3
The player who is in the role of Player A after the lottery receives an additional 100 points. **Just as in Part 1, Player A can distribute these 100 points between herself and Player B.** Contrary to Part 1, you already know at this point whether you are the sending (A) or receiving (B) player. Player B receives the amount given by Player A, who keeps the remainder of the 100 points.

Your points from Part 3 are thus calculated as follows:
Points for Player A = 100 - points sent to Player B
Points for Player B = 0 + points sent from Player A

Payout
We kindly ask you to fill out an additional questionnaire. Your answer here do not influence your monetary payout. Afterwards, you will be informed of your earnings for today. Your payout will consist of the points you earned during the Parts 1,2, and 3 as well as the points from a randomly selected guessing question.

Points from Part 1
\+ Points from Part 2
\+ Points from Part 3
\+ Points from a guessing question
= Total points

Total points ÷ 18 = Payout in €

Player B is informed on his computer screen how many points Player A has allocated to him and in which cubicle Player A is sitting. Afterwards, all participants open the curtains of their cubicles. Furthermore, all participants who were in the role of player A stand up and step forward to the floor marking in front of their cubicle.

For each player A, the number of the cubicle is read out and it is announced how she has divided the 100 points in Part 3 between herself and player B. No further information, e.g., the name, will be made public. Player B will not stand up and Player A will not be able to identify her partner. Every participant today will thus be informed about every Player A's sharing decision.

Afterwards, we will call you one by one to receive the full payout. Only you and one lab assistant will know the full amount of your earnings today.

Chapter 4

The Cost of Being First: Self-Image Concerns and the First-Offer Dilemma in Bargaining

Authors: Christos Litsios

Abstract

This paper addresses the first-offer dilemma in bargaining by investigating the role of self-image concerns. Our online experiment builds upon the preference elicitation strategy of Litsios and Schories (2024). A second-price auction sells a probability advantage in a lottery, allocating the dictator role in an ensuing dictator game. We use the dictator game to isolate the behavioral motive from other strategic considerations. We presume that subjects face a trade-off when "making the first move", because the offer conveys a signal which can stimulate self-image concerns. Individuals concerned about their self-image experience psychological costs from behavior that differs significantly from their perception of the relevant social norm. Anticipating the psychological costs, they restrain themselves from selfish behavior, thereby weakening their first-mover advantage. We manipulate self-image concerns by offering two avoidance strategies and observe how this affects bidding behavior as a proxy for the first-mover preference. Under belief manipulation, subjects adjust their perception of the relevant social norm Bicchieri et al. (2023). Under strategic ignorance, subjects deliberately remain ignorant about their recipient's payoff consequences Dana et al. (2007). We find evidence that subjects use belief manipulation to behave selfishly as dictators, and mild evidence that this translates into their first-mover preference. Although our design fails to isolate strategic ignorance, we observe how the combination of pessimistic beliefs and ignorance dramatically deteriorates prosocial behavior. In conclusion, our experiment highlights the role of self-image concerns via beliefs in the first-offer dilemma.

Keywords: Self-Image, Dictator Game, Auction, First-Offer Dilemma

JEL Classification: A13, C91, D44, D64, D91

4.1 Introduction

Bargaining is a phenomenon at the heart of economics. From the simplest form of a two-person trade to complex multiparty contractual arrangements, bargaining is the primary mechanism for allocating resources in economic activity. Thus, it has received considerable attention in economics, business, negotiation, and psychological research. This paper focuses on the very first step of the bargaining process - the first offer. We attend to what has become known in the literature as the first-offer dilemma. The dilemma lies in a trade-off between the benefits and costs of placing the first offer.

Practitioners and experts in negotiation and bargaining have postulated *to never make the first offer unless you have to* (Dell and Boswell (2009), McCormack (2022)). In their argument, the first offer can serve as a signal, revealing valuable information that would otherwise be lost. Researchers, on the other hand, advocate for the value of the first-mover role. Economists, dating back to Rubinstein (1982), note that first-movers can structure the bargaining process with their first offer. Ochs and Roth (1989) provide empirical evidence on this first-mover advantage. Psychologists emphasize the *anchoring effect* of the first offer (Galinsky and Mussweiler (2001), Orr and Guthrie (2005), Gunia et al. (2013)). They argue that the first offer affects the individual's cognitive process of information acquisition and sets a focal point for the final agreement.
More recent research has aimed at uniting the contradictory anecdotal and empirical evidence. Cotter and Henley Jr (2008) point out that the existence of a first-mover advantage depends on the experience of the opponent. Making the first move against an experienced opponent more often results in less favorable outcomes and appears as a disadvantage. Loschelder et al. (2014) speak of the *practitioner-researcher paradox* and identify a first-mover disadvantage to exist whenever the first-mover reveals compatible preferences with their offer. The offer serves as a valuable signal, and the recipient can use the information to their advantage. Loschelder et al. (2016) propose the *Information-Anchoring Model of First Offers* for multi-issue bargaining and explain under what conditions the first offer works towards the sender's advantage and disadvantage. In simple distributive issues, the first offer serves as an anchor and works to the sender's advantage. Under integrative issues, the first offer can transmit valuable insights into the sender's preferences. A selfish recipient can pretend to concede on a compatible issue and gain on the distributive ones. Thereby, the signalling property of the first offer becomes a disadvantage. Osório (2020) is the first in economics to address the first-offer dilemma in a game-theoretic model. They extend the Rubinstein (1982) alternating-offers model to incorporate incomplete information about the players' patience types. The model explains a first- and second-mover advantage depending on different parameter constellations. Most interestingly, for a specific range, they find a second-mover advantage as the first offer signals the sender's type.

We approach the topic from a behavioral economics perspective and ask whether the self-image concerns affect the first-mover preference. Litsios and Schories (2024) were the first to investigate the role of image concerns on the first offer. They designed a novel preference-elicitation mechanism and tested whether social image concerns affect first-mover preferences. We build on their mechanism to investigate the role of self-image concerns on the first-mover preference. We conduct an online experiment in which we auction the dictator role in a second-price auction to measure the subjects' first-mover preferences. We control self-image concerns by offering participants two potential self-image avoidance strategies and measuring their impact on the first-mover preference between treatments. The two strategies we use are belief manipulation (Bénabou and Tirole (2016) and Gino et al. (2016)) and strategic ignorance (Dana et al. (2007)).

Belief manipulation enables individuals to behave selfishly while maintaining a positive self-image by strategically selecting beliefs about appropriate behavior. Any image concern must include a perception of what behavior is desirable and, therefore, a standard against which one's behavior is measured. Behavior close to that threshold translates into a positive image. The desired behavior in a given context is defined by an individual's perception of the relevant social norm. Bicchieri (2016) postulate that this perception is built on one's normative and empirical expectation of a relevant reference group. These expectations are beliefs about what the reference group thinks one ought to do and what they actually do. Before the main part of the experiment, we elicit both beliefs in an incentivized task similar to the well-known coordination task in Krupka and Weber (2013). Our treatment varies in the timing of information provision about the subsequent course of events. Dependent on whether subjects are aware of their own participation in the main part of the experiment, belief manipulation is possible or not. The procedure of withholding information before the coordination task has been applied by Bicchieri et al. (2023), where subjects are observed to distort their beliefs, especially their empirical expectations, in a self-serving manner. Strategic ignorance allows individuals to behave selfishly while maintaining a positive self-image by deliberately ignoring the consequences of their actions. We apply a treatment that offers subjects the opportunity to remain ignorant about the outcome of their dictator choice for the recipient. Before the dictator decision, one of two matrices is randomly selected to be payoff-relevant. In one matrix, the dictator's and the recipient's incentives are aligned, while in the other, they are not. The subjects can then choose whether to reveal the active matrix or remain ignorant about the recipient's payoff consequences. This design was first used by Dana et al. (2007) to test whether prosocial behavior results from a stable social preference or is more context dependent. They found increased selfish behavior when there was an opportunity for strategic ignorance.

Our study contributes to the economic literature in several ways. First, we extend research on prosocial behavior by shedding light on the role of self-image concerns in distributional decisions. Our findings support Dana et al. (2007) and others, who argue that prosocial behavior is context-dependent and driven by a more complex psychological motive than a simple preference for fairness. Second, we apply this psychological motive to the archetype of economic interaction – bargaining and negotiation. Especially, we examine another behavioral layer of the first-offer dilemma. Third, by applying the avoidance strategies of belief manipulation and strategic ignorance, we further investigate how self-image concerns can be manipulated, while forth, we deepen our understanding of the connection between self-image concerns and social norms. Finally, our research adds to our understanding of the benefits and shortcomings of online experiments.

Our results give rather weak or no support for our hypothesis about the importance of self-image concerns for the first-mover preference. On the level of the dictator game, our two manipulation strategies show an impact. First, we find mild evidence that belief manipulation increases the first-mover preference. Subjects' beliefs shift in a pessimistic direction, and we observe more selfish behavior. Average auction bids do not differ significantly across treatments, and despite an increased number of selfish subjects, bid distributions remain unchanged. Average bids and distributions of selfish subjects do not differ between treatments. This suggests that belief manipulators have preferences similar to those of purely selfish subjects; however, given the small sample sizes, the increase in the number of selfish subjects does not translate into a significant shift in auction bids in the global comparison of treatments. Second, our attempt to isolate the effect of strategic ignorance was unsuccessful because a design flaw introduced additional variation between the treatments, making it difficult to attribute changes in behavior solely to strategic ignorance. Anyhow, the combination of effects in the strategic ignorance treatments almost entirely deteriorates pro-social behavior and highlights the importance of self-image concerns.

The paper proceeds as follows. In the second chapter, we will summarize relevant literature from primarily behavioral economics and provide a theoretical background for our experiment. The third chapter introduces the experimental design. Chapter four outlines our hypotheses. In chapter five, we present our data analysis. The final chapter concludes with a short discussion.

4.2 Literature & Theoretical Background

Economic research on bargaining has produced insights into the first-offer dilemma by focusing on strategic aspects (see, e.g., Rubinstein (1982), Ochs and Roth (1989), and Osório (2020)). This paper approaches the dilemma from a behavioral economics perspective and examines how self-image concerns shape the first-mover preference. We isolate the effect of these non-financial motives from other strategic considerations by focusing on the first stage of the alternating-offer structure. The dictator game gives all the bargaining power to the first mover, and the second mover becomes a passive recipient. Suppose we observe a change in bidding behavior as we manipulate self-image concerns, even when the dictator has all the bargaining power. In that case, these psychological costs must be considered in the first-offer dilemma.

Since its introduction by Kahneman et al. (1986), the dictator game has been widely used to understand the driving motives in decision-making. Standard game theory predicts that the dictator will take advantage of their position, but there is plenty of empirical evidence that shows subjects sharing non-zero amounts with their recipients (e.g., Engel (2011), Camerer (2011)). In a meta-study, Engel (2011) summarizes that dictators, on average, share 28.35% of the surplus. About 36.11% give nothing, and 16.74% decide for the equal split. Influential theories explained pro-social behavior by an intrinsic preference for altruism or fairness in the utility function (e.g. Fehr and Schmidt (1999), (Bolton and Ockenfels, 2000)). However, earlier experimental studies by Hoffman et al. (1996) and Bohnet and Frey (1999) already cast doubt on the idea that an intrinsic preference for simple outcome equality alone can account for the phenomenon. Instead, alongside other factors, they observed a negative relationship between prosocial behavior and the dictator's degree of anonymity. Similar doubts were raised by Dana et al. (2006). In their experiment, dictators abandon a surplus to silently exit the game and walk away without the recipient's notice. Dictators were offered to either stay and play a regular dictator game allocating 10 USD, or leave without anyone's notice. Leaving, they got 9 USD. Payoff dominance prescribes staying and opting for 10 USD, or choosing 9 USD and equalizing payoffs. Subjects frequently chose to silently escape. Additionally, subjects participated in a 'private' dictator game, in which recipients never knew the money they received came from a dictator game. There, only a few subjects exited the game. The authors conclude that prosocial behavior is rather driven by a need to fulfill social expectations than by an honest interest in others' welfare. Andreoni and Bernheim (2009) design social image concerns in a signaling model of the dictator game. They extend their model from simple compliance to an equality norm by social image concerns. The image concern is modeled as an outside observer's conditional expectation of the norm-compliance preference given the dictator's choice. The choice functions as a signal. The signaling motive leads to higher offers for some individuals and eventually explains pooling behavior of different preference types at the prevailing norm. An experiment supports their model's predictions. Litsios and Schories (2024) are the first to investigate the role of social image concerns on the first-mover preference in bargaining. They design an innovative experiment and show that ex post publication of the dictator's identity affects not only the dictator's giving but also the subjects' dictator preferences. They formalize a game-theoretic model based on Andreoni and Bernheim (2009). The model explains the underlying mechanics between image concerns and first-mover preferences. In their experiment, they match two subjects and have them compete in a second-price auction for the dictator role. Their public treatment reveals the dictator's identity ex post, thereby stimulating social image concerns. Behavior that deviates from social norms can lead to a negative image and incur psychological costs. In the public treatment, image-concerned subjects adjust their behavior in anticipation of the costs and accordingly bid less in the second-price auction.

Auctioning the dictator role is a novel strategy to measure first-mover preferences. There have been studies that sell participation in the ultimatum game in predefined roles via a second-price auction in Güth and Tietz (1986) and via an English clock auction in Shachat and Swarthout (2013). Subjects were randomly assigned to either a proposer or a responder role and then competed in the corresponding auction. Both studies find higher bids for the proposer role. Anyhow, the studies concern the ultimatum game, and their experimental strategy aims to reveal preferences for participating in the game in either role. It does not directly address our research question. The choice of a second-price auction to elicit first-mover preferences is founded on its theoretical equilibrium properties. A weakly dominant strategy in the simple static game version of the auction is to bid one's true valuation of the good to be sold. By resorting to their weakly dominant strategy, subjects can bid their true valuation without accounting for their unrefined beliefs, risking overbidding or losing their stake on terms they could have prevented. This rationale allows interpreting changes in bidding behavior across treatments as an indication of changes in first-mover preference.

We use a similar experimental setup to Litsios and Schories (2024). A dictator game follows a second-price auction to investigate the impact of self-image concerns on dictator-giving and the first-mover preference. Although social image concerns have been shown to instigate prosocial behavior when the individual's identity is revealed, it remains to be investigated what causes the residual prosocial behavior under anonymity. We propose that for self-image concerns, a similar signaling rationale applies as described in the model by Andreoni and Bernheim (2009), only now, signaling is directed towards the individual itself. The behavior signals their inherent preference, and specific individuals are concerned about their preference type - their idea of themselves, their identity. Influential studies by Bodner and Prelec (2003) and Bénabou and Tirole (2006) model self-image concerns in a very similar fashion to social image concerns in Andreoni and Bernheim (2009). The difference between social and self-image concerns is that the individual themselves has an interest in observing their own behavior. The motivation relies on a central assumption about the individual's memory. The authors of the former two papers refer to an argument made by Bénabou and Tirole (2004). They propose that individuals have imperfect recall and can remember their actions and not their inherent preferences. Therefore, they infer their preference type from their actions. The idea is rooted in self-deception theory by Bem (1972).

Research on motivated beliefs concerns how individuals strategically choose self-serving beliefs to justify their selfish behavior to themselves (see, e.g., Bénabou and Tirole (2016) and Gino et al. (2016)). A commonly cited example is given in Babcock et al. (1995). In an experiment, subjects are confronted with a fictional legal dispute over a car accident with ensuing negotiations for settlement. Respectively, two subjects are allocated the roles of plaintiff and defendant. After reading the case materials, they are left to evaluate the amount a judge would award the plaintiff if negotiations fail. After that assessment, subjects engage in actual settlement negotiations. In one treatment, the roles are allocated after both subjects have read the case material and made their assessments under this veil of ignorance. The main result of the study is that when roles are assigned before the assessment, subjects' beliefs about the judge's decision are distorted in their favor and therefore diverge from their opponents' assessments. On the contrary, when the assessment is made under the veil of ignorance, both assessments coincide, and negotiations are more successful. Di Tella et al. (2015) is even closer to our research question. They employ a variation of the dictator game to test for strategic belief manipulation. They are interested in beliefs about the recipients' trustworthiness and how beliefs affect dictator offers. They let subjects play a corruption game. Two parties play a limited-offer version of the dictator game. The surplus is 20 tokens, initially divided into two piles of 10 tokens each. The dictator's decision is whether to scrape tokens from the recipient's pile or not. Simultaneously with the dictator, the recipient can take a side payment, which shrinks the entire surplus. Dictators are assigned randomly to one of two conditions. The conditions differ in the amount that the dictators can take from the recipient's pile. In *able=8* (*able=2*) dictators can take up to 8 (2) tokens. Recipients are unaware of the limited offer options imposed by the treatments. Before the corruption game, dictators' beliefs about the proportion of recipients opting for the side payment are elicited by an incentivized task. In the *able=8* treatment, dictators expect recipients to take the side payment with a 20 percent higher probability than in the *able=2* treatment. Dictators considerably distort their beliefs in a self-serving manner. More pessimistic beliefs about the recipient's behavior are associated with dictators appropriating, on average, 2.5 additional tokens. Maintaining the proper belief appears to be a strategic choice to justify more selfish behavior in the future. Gangadharan et al. (2024) evaluate different belief elicitation methods when incentives between a belief elicitation task and a donation decision compete with each other. Their setup invites belief manipulation in an elicitation task to garner a positive self-image while still behaving rather selfishly in the donation decision. They propose a simple theoretical framework, also based on Bodner and Prelec (2003) and Bénabou and Tirole (2006), that ascribes self-serving belief manipulation to self-image concerns. Individuals face a tradeoff between financial payoffs and psychological costs. In the donation decision, subjects can allocate a small part of their endowment to a charity. In the elicitation task, subjects are asked how many participants decided to give money. The authors use three different mechanisms - one without incentives, one with incentives where a correct guess leads to a payoff, and third, a Karni mechanism (Karni (2009)) where the chance of receiving a payoff increases in the accuracy of the beliefs. As expected from the model's predictions, they find that non-donor beliefs depend on the elicitation mechanism, with *no incentives* producing the greatest belief distortion, followed by *incentives* and, finally, the Karni mechanism. As we want individuals to take advantage of belief manipulation to manipulate self-image concerns, we use an elicitation mechanism with simple incentives.

In most bargaining situations, no law governs which distribution is the right one to implement. Instead, it is the framing of the situation and a social norm that prescribes which allocation makes an action appear appropriate. For example, while in most contexts the prevailing social norm in the dictator game would prescribe an equal split, as in the much-cited study by Krupka and Weber (2013), there are examples, such as Dufwenberg and Muren (2006), that point in the opposite direction. There, subjects were recruited in an economics lecture, and once the dictator's behavior was revealed to their peers, more selfishness emerged. It seems reasonable to assume that the norms among economics students differ from those of the general public. Therefore, what constitutes appropriate behavior and results in a positive image may vary depending on perceptions of the prevailing norm. Even though self-image concerns are intrapersonal phenomena, they also need to be informed by a normative standard. What an individual understands as appropriate behavior is therefore closely connected to their understanding of who they are and how they connect to their social environment - their identity. Akerlof and Kranton (2000) discuss identity and how it shapes economic behavior. They use the term 'prescriptions' interchangeably with 'norms', as rules of behavior that manifest affiliation with a certain group, and adherence to which can thereby define one's identity. The two concepts, self-image concerns and identity, are closely connected. By behaving in accordance with our idea of how a member of a specific group ought to behave, we claim our identity. In this way, self-image concerns depend on the active social norm in the relevant peer group. We follow a conceptual understanding of social norms defined by Bicchieri (2016). In short, social norms are behavioral rules governing certain decisions. Adherence to this rule is conditional on an individual's empirical and normative expectations about their relevant peer group. Empirical expectations express how far an individual expects others in their peer group to adhere to the behavioral rule. Normative expectations express how far the individual expects others in their peer group to understand the behavioral rule as valid and to punish others who do not adhere to it. In an experimental study by Bicchieri et al. (2023), the authors investigate self-serving belief manipulation in honesty decisions. They investigate the impact both kinds of expectations have on honesty in a version of the die-rolling game. Before subjects engage in the main game, they complete an incentivized belief-elicitation task about either their empirical or normative expectations. The task elicits subjects' expectations about

the majority of subjects in a similar study on the die-rolling game. The authors change the provision of information on one treatment dimension and the retrieval of empirical or normative expectations on the other. On the information dimension, they vary the timing of informing their subjects about participating in the die-rolling game. In one treatment, subjects first complete the belief elicitation task and then are informed that they will participate in the die-rolling game. In the other treatment, participants are informed about the die-rolling task before they play the belief elicitation task. Thereby, the authors vary the subject's possibility to manipulate their own beliefs in a self-serving way. Only if they are informed about their participation before the belief elicitation task are they in a position to strategically choose their beliefs. In the other treatment, subjects first commit to a belief, having in mind maximizing their payoff in the current task, and then learn about their further participation. The authors find evidence for belief manipulation in empirical expectations and its impact on the honesty decision. Subjects seem to defuse the social norm for honesty by successfully convincing themselves that "the majority lies" and thereby it is acceptable for them to lie too.
An earlier study by Bicchieri and Xiao (2009) investigated the efficacy of both expectations on the subject's norm adherence in a dictator game. They found similar evidence supporting the significance of empirical expectations in decision-making. Holding the empirical belief that "if others do not follow, I do not need to follow" legitimizes selfish behavior and appears to be used to justify one's advantage. We use a very similar experimental set-up to Bicchieri et al. (2023) to manipulate self-image concerns.

As a second alternative approach to manipulating self-image concerns, we apply the strategy described by Dana et al. (2007). The authors conduct an experiment introducing different forms of 'moral wiggle room' to test the origin of prosocial behavior in the dictator game. They construct three treatments of a binary choice dictator game. In various ways, they invite uncertainty to obfuscate transparency into the consequences of dictators' actions. All three treatments result in more selfish behavior in the dictator game compared to a baseline treatment. Conclusively, the authors doubt that generous outcomes in dictator games are evidence of an inherent fairness preference. Instead, certain situational factors encourage prosocial behavior by mediating alternative motives. Especially their first treatment (*hidden information*) offers valuable insights for our considerations. Here, the dictator can choose to stay ignorant about the payoff consequences for the recipient. If possible, many dictators prefer to ignore information on the recipient's payoff consequences and increasingly choose the selfish option. Namely, 44% of dictators preferred to remain ignorant, and among them, 86% chose the selfish option. Obscuring the recipient's payoff consequences encourages selfish behavior by fostering what the literature calls strategic or willful ignorance. Grossman and Van Der Weele (2017) pick up on the empirical observations made by Dana et al. (2007). They develop a model based on a similar environment to that of Dana et al. (2007), which explains how strategic ignorance is used to circumvent social image concerns and to resort to more selfish behavior. They are especially concerned with explaining how deliberately choosing ignorance remains a valid strategy for preventing a negative self-image. They propose a signaling model and conduct experiments to explore the connection. Theoretically, they prove the existence of an "ignorance equilibrium", in which self-image-concerned agents with low intrinsic altruism strictly prefer to stay ignorant. Namely, people prefer to avoid situations in which high financial incentives coincide with the potential for high signaling costs. They cite an example in the domain of climate change, previously stated by Norgaard (2006) - people prefer to retain some ignorance to consume with a clear conscience. In conclusion, they argue that their theory and empirical evidence provide probable cause to state that "self-signaling is an important driver of behavior in social situations" and that strategic ignorance offers a way to evade costly investments in a positive self-image.

4.3 Experimental Design

[Experimental Design]

The experiment is programmed in oTree (Chen et al., 2016), conducted on a *Heroku* server, and participants are recruited via the online research platform *Prolific*. We employ a 2 × 2 between-subjects design with four treatments. Each participant is randomly assigned to one condition. Table 4.1 depicts the four different treatments within the two-dimensional condition space.

Table 4.1: **2×2 Treatment Design.** The treatments differ on two dimensions. *Visibility* indicates whether subjects observe their recipient's payoff consequences behind their dictator's decision in the *open* treatment or whether those consequences are initially hidden and they can choose to uncover them in the *hidden* treatment. *Information* differentiates whether subjects know about their participation in the main game at the time of the expectation elicitation task in the *before* treatment or whether they are unaware of their participation in the *after* treatment.

		Visibility	
		open	*hidden*
Information	*before*	T1	T3
	after	T2	T4

On the visibility dimension, we differentiate between the *open* and *hidden* conditions. In the *open* treatments, subjects are fully informed about the consequences of their decision as dictators. In the *hidden* treatments, subjects can either remain ignorant of their recipient's payoff consequences or reveal them before the decision. By equal chance, one of two different payoff matrices is active. Choosing ignorance leaves the dictator unaware of the active matrix and therefore of the recipient's payoff. In both matrices, the same dictator choice leads to the same dictator payoff, while the recipient's payoffs are inverted between the two matrices. The visibility dimension regulates the potential for strategic ignorance. We will discuss the details later in this section. On the information dimension, we differentiate between the *before* and *after* conditions. In the first part of the experiment, subjects complete an incentivized belief elicitation task[1]. In the *before* treatments, subjects know, ahead of their decisions in the elicitation task, the remaining course of the experiment. In the *after* treatments at the time of the decisions in the elicitation task, subjects are unaware of the remaining course of the experiment. The *after* treatment leaves them unaware of their participation in the second-price auction and the dictator game. The information dimension, therefore, regulates the potential for belief manipulation. In the following, we will further explain the timing of the experiment, illustrated in figure 4.1. Afterward, we will discuss the single components in greater detail.

Timing

[1]For simplicity, we continue to refer to it as a *belief* elicitation task. More thoroughly, we investigate normative and empirical expectations in coherence with Bicchieri (2016).

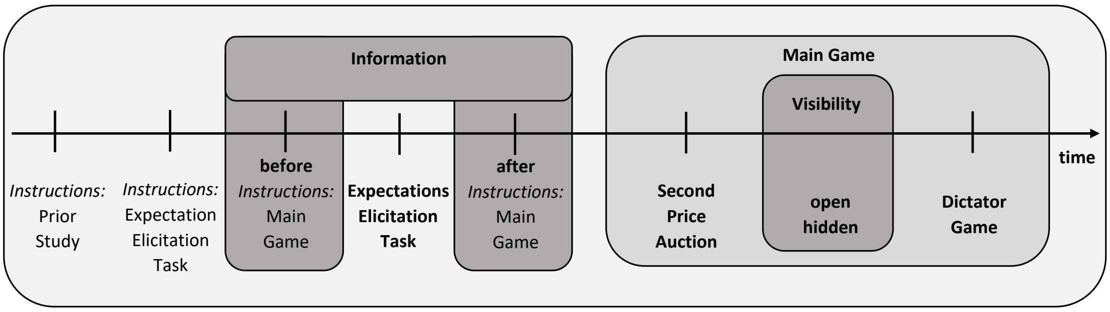


Figure 4.1: Timing of the Experiment.

In all four treatments, the experiment begins by introducing the subjects to a prior study that uses a reduced form of the actual experiment. In the prior study, subjects played a similar belief elicitation task, followed by the same dictator game without the second-price auction. The prior study included only the two treatments along the visibility dimension, *open* and *hidden*. We chose to run this simplified version before the actual experiment to gather data on expectations and behavior to incentivize the belief elicitation task. We will further explain this choice when introducing the belief elicitation task.
Subsequently, the subjects receive the instructions for the belief elicitation task. After the instructions, the subjects need to answer a set of comprehension questions. The experiment does not continue before all questions are answered correctly. In case of an incorrect answer, we present some explanation and the correct answer.
Depending on the information condition, subjects are either presented with instructions for the main game or are directly taken to the elicitation task. In the *before* treatment, subjects get the instructions before the elicitation task and are explicitly informed about their participation in the main game. In the *after* treatment, subjects receive the full instructions only after the elicitation task, and there is no indication that the experiment continues beyond the elicitation task. In each case, after the instructions for the main game, another set of comprehension questions awaits the subjects. Again, all questions need to be answered correctly to continue the experiment.
The main game comprises the second-price auction and the dictator game. The instructions for the main game include explanations for both stages and information about the active visibility condition. In the *open* treatment, there is common knowledge of the payoff matrix in the dictator game. In the *hidden* treatment, initially, the recipient's payoff is concealed. Covertly, one of two payoff matrices is active. Each subject faces an additional decision: whether to uncover the active matrix and thereby become informed about their recipient's payoff consequences, or remain ignorant. This decision is made after the second-price auction. At the time of the auction, subjects do not know which matrix is active in the *hidden* treatment and are fully aware of the effective payoff matrix in the *open* treatment. In the auction, subjects are paired in anonymous couples and compete for a probability advantage of becoming the dictator in the ensuing dictator game. After the auction, a lottery determines the dictator within each couple. Subjects only learn about their role and payments at the very end of the experiment. In the *open* treatments, both subjects directly proceed to the dictator decision. In the *hidden* treatments, the subjects first decide whether to reveal the active matrix or remain ignorant before proceeding to the dictator decision. With the subjects only being informed about their respective roles after the entire experiment, the potential reveal and dictator decisions are made by both subjects through the strategy method. After the dictator decision, subjects learn about their role and payoff, and the experiment ends with a short survey.

Information Condition

The design choice of the information dimension is inspired by Bicchieri et al. (2023). The information dimension varies between the *before* and *after* treatments regarding the moment subjects are informed of their participation in the main game. In the *after* treatment, subjects approach the elicitation task without any hint that the experiment will continue afterward. They receive information about the main game only after they make their decisions in the elicitation task. We can assume that those decisions are made independently of future interests in the main game. In the *before* treatment, subjects are informed of their participation in the main game before engaging in the elicitation task. This way, the entire experiment is laid out in front of them, and they can make an informed choice in correspondence with their future interests.
The theory of self-image holds that any action can only be evaluated against a specific benchmark. The benchmark would be the social norm that applies to a given individual. According to the definition of social norms by Bicchieri (2016), they rely on an individual's normative and empirical expectations about their relevant peer group. The concept of belief manipulation claims that individuals adjust their expectations to justify certain behavior, for the sake of their self-image. By the information dimension, we intend to turn belief manipulation on and off. Under the *before* condition, subjects are aware of the entire experiment. By means of backward induction, they understand the value of advantageous expectations in maintaining a positive self-image. Under the *after* condition, subjects are unaware of the main game, and their expectation formation is clear of any such considerations about the future. We assume that the *before* treatment offers a way to disable self-image concerns by belief manipulation. Therefore, we can measure the impact of self-image concerns between the two conditions.

Belief Elicitation Task

The task elicits normative and empirical expectations, which together constitute a social norm as defined in Bicchieri (2016). It follows a similar logic to the norm elicitation procedure first used by Krupka and Weber (2013), but in its specifics closely resembles the elicitation task in Bicchieri et al. (2023). We ask subjects about their expectation of other subjects' opinions about 'what ought to be done' (normative expectations) and others' 'actual behavior' (empirical expectations) in the dictator game of the prior study[2]. Subjects receive £0.1 in monetary compensation for each correct guess. The entire task is presented to subjects on one webpage. We use the phrasing *opinion* and *behavior* assessment, and in each case, ask which of two mutually exclusive statements is true. In the opinion assessment, evaluating subjects' normative expectations, subjects choose between "In the earlier study, most people said it is OK to choose A(B).". In the behavior assessment evaluating empirical expectations, subjects choose between "In the earlier study, most people chose A(B)". A and B are the two dictator choices, which will be further explained in the subsection on the dictator game. Both assessments relate to the majority's choice and do not ask for a specific estimate of the distribution of choices. On the one hand, this simplifies the task; on the other hand, it limits quantitative analysis to a simple comparison of frequencies between treatments. Subjects learn the correct answer and their payment only after the entire experiment.

[2]The prior study was needed to circumvent a problem posed by the information condition. In the *after* treatment, during the elicitation task, subjects are unaware of their participation in the main game. We want to elicit empirical expectations, i.e., expectations about actual behavior in the dictator game. Informing subjects about the conditioning of payments on behavior in the current study would have posed the problem of changing the information structure and dismantling the impact of the *after* treatment.

Main Game

The main game of the experiment contains the second-price auction and the dictator game. The major difference in the visibility dimension plays out at this stage. The payoff structure in the dictator game differs between *open* and *hidden* conditions. Under the *hidden* condition, subjects can choose to reveal the recipient's payoff or remain ignorant.

Second-Price Auction

Two subjects are anonymously paired and compete in a second-price auction. The auction winner is granted a probability advantage over their opponent in a lottery that determines the dictator in the subsequent dictator game. More precisely, the auction winner becomes dictator by 90% chance and the loser by 10%[3]. As is common in a second-price auction, the winner pays the loser's bid, and the loser pays nothing. We decided on a second-price auction because of a useful equilibrium property in its static game representation. In a weakly dominant strategy, both players reveal their true valuation for the auctioned good without the need to consider their opponent's type. Irrespective of players' beliefs about their opponents' type distribution, choosing their true valuation is a safe bet, as they never pay too much to win and are insured against paying more than they value. On this theoretical basis, we compare the bids in the second-price auction between treatments to estimate systematic differences in subjects' valuation - their preference for the dictator role[4].
Each subject receives an initial endowment of £1.8 at the beginning of the experiment without any further explanation. Subjects can place their bid from this endowment. A bid can be any discrete £amount between zero and £1.80. Subjects are reminded of their guesses in the elicitation task and the further course of action on the auction screen. In the case of the *hidden* condition, they are reminded of the revelation decision following the auction. When subjects place their bid in the auction, they are unaware of the active matrix. And finally, they are reminded of what is at stake in the auction, namely, the probability advantage in the role lottery. Subjects are asked "How much of your endowment would you like to bid?". Consequently, the roles are assigned within the couples but kept secret until the end of the experiment. Subjects only learn about their role and the associated cost of achieving it after the decision in the dictator game for two reasons. First, assuming the dictator role by an auction potentially entails entitlement effects as described, for example, in Hoffman et al. (1996). When subjects understand that they have earned their role as dictators, they behave more selfishly, driven by a meritocratic understanding. Our setup leaves subjects in the dark and obfuscates the role assignment process. Second, we opted for the strategy method in the dictator game to increase the number of observations. Therefore, they need to be ignorant about their role at the time of the decision.

Dictator Game

In the binary dictator game, subjects are unaware of their own role. We apply the strategy

[3] In case of a draw, chances are even and the winner pays their bid.

[4] More precisely, we measure the financial value of the preference for a probability advantage. One shortcoming in this method is that we do not further control for diversity in risk preferences. However, since the difference in expected payoffs is fairly low, we assume risk neutrality.

method, and both subjects make a hypothetical dictator decision[5]. Subjects choose between two alternative allocations, namely $A = (X_A, Y_B)$ and $B = (X_B, Y_B)$. X_i with $i = A, B$ is the monetary payoff for the dictator and Y_i with $i = A, B$ is the monetary payoff for their respective recipient. Organizing the two allocations into a 2x2 matrix yields the payoff matrix. The payoff matrix differs on the visibility dimension between the *open* and *hidden* conditions described in the following subsection.
To avoid framing effects, we give the two roles generic names. Player X stands for the dictator, and Player Y for the recipient. On the screen of the dictator's decision, subjects are once more reminded of their guesses in the elicitation task and the relevant payoff matrices. Then they are asked " Which allocation do you choose, and therefore will be executed *if* you become Player X?".

Visibility Condition

Under the *open* condition, subjects are fully informed about the active payoff matrix in the introduction to the main game. We name it matrix 1 and depict it in figure 4.2.

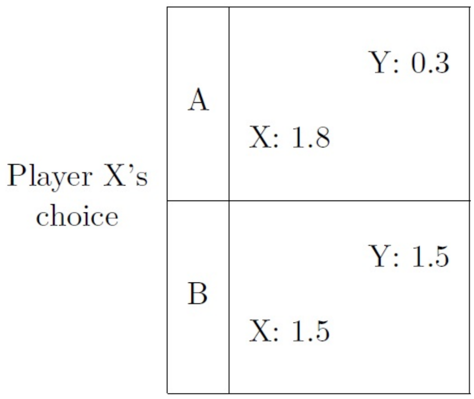


Figure 4.2: **Payoff Matrix 1 under *open* Condition.** This table presents the payoff matrix under the *open* condition. Player X, the dictator, has two choices. They either choose allocation A or B. The upper cell lists the payoffs for both players under allocation A. Player X gets £1.8 and Player Y, the recipient, receives £0.3. The lower cell lists the payoffs for both players under allocation B. Player X gets £1.5 and Player Y receives £1.5 as well.

In the *open* treatments, every subject knows about the consequences of the two allocations, A and B. When Player X, the dictator, decides on allocation A, they get £1.8, and Player Y, the recipient, receives £0.3. When Player X decides on allocation B, they get £1.5, and Player Y receives £1.5 as well. Allocation B presents an equitable outcome, whereas allocation A gives the dictator an individual advantage and neglects the recipient. We call allocation A the selfish choice and B the equitable choice. The initial endowment and the dictator's payoff from the selfish choice are chosen to match each other. Thereby, subjects can bid at most exactly the

[5]The strategy method allows us to collect decisions from both subjects and increases the number of observations. Brandts and Charness (2011) find no clear difference in results between strategy and direct-response method.

winnings from the selfish dictator's choice in the auction.

Under the *hidden* condition, subjects are initially ignorant about the recipient's payoff consequences Y_i from the two choices A and B. In the main game instructions, they learn that either matrix can be active. One is matrix 1, the same as in the *open* treatments, and the other is matrix 2. Both are illustrated in figure 4.3.

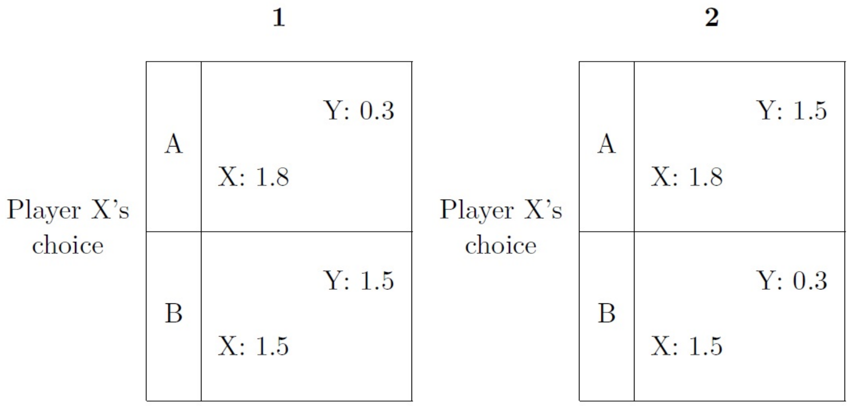


Figure 4.3: **Payoff Matrix 1 & 2 under *hidden* Condition.** The left panel presents payoff matrix 1 known from the *open* treatments. The right panel presents payoff matrix 2. Player X, the dictator, has two choices. They either choose allocation A or B. The upper cell lists the payoffs for both players under allocation A. Player X gets £1.8 and Player Y, the recipient, receives £1.5. The lower cell lists the payoffs for both players under allocation B. Player X gets £1.5 and Player Y receives £0.3. In both matrices, Player X's payoffs are the same for the two choices. Player Y's payoffs are switched.

The left panel presents payoff matrix 1, the same as in the *open* treatments. The right panel presents payoff matrix 2. In both matrices, Player X gets £1.8 from choosing allocation A and £1.5 from choosing B. The difference between the two matrices lies in Player Y's payoffs. The recipient's payoff consequences from the two choices are exactly switched between the two matrices. In matrix 1, we labeled the two allocations as A (selfish) and B (equitable). In matrix 2, these terms no longer fit. Indeed, allocation A gives Player X an advantage over B, but now also makes Player Y better off by assigning them £1.5. Allocation B makes both worse off and assigns £0.3 to the recipient. There is no apparent reason why Player X should ever choose B whenever matrix 2 is active. With subjects fully informed about which matrix is active, their choice is between weighing financial motives against image concerns in matrix 1 and, most likely, a simple choice for A in matrix 2. Instead, at the outset of the *hidden* treatments, the active matrix is concealed. In the background, a 50:50 random draw determines which of the two matrices is active.

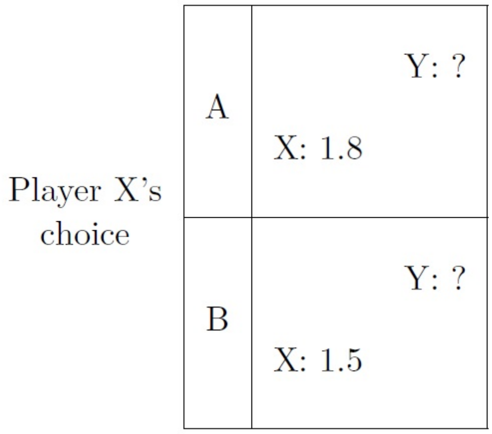


Figure 4.4: **Initial Payoff Matrix under *hidden* Condition.** This table presents the initial payoff matrix under the *hidden* condition. Player X, the dictator, has two choices. Either they choose allocation A or B. The upper cell lists the payoff for Player X under allocation A - £1.8, while the payoff for Player Y remains unknown. The lower cell lists the payoff for Player X under allocation B - £1.5, while the payoff for Player Y remains unknown.

In the *hidden* treatments, at the time of the dictator decision, subjects see table 4.4. The payoffs for Player X, the dictator, are uncovered and the same as in matrix 1. Instead of £amounts, subjects find question marks for Player Y's payoffs. Before they make their dictator decision on which allocation to enforce, they can reveal the active matrix at no cost by clicking a button. If they decide to reveal the game, either matrix 1 or matrix 2 appears, and they finalize their dictator decision. If they instead refrain from revealing the game, they make their dictator decision directly. The *hidden* treatments open up the possibility of strategic ignorance. Uncertainty about the recipient's payoff consequences provides a preferential rationale for opting for allocation A. A dictator who stays ignorant about which option is active can soothe herself into believing that, after all, choosing A is not such a clear signal of selfishness. Ignorant subjects remain so even after the experiment.

4.4 Hypotheses

In our first approach to controlling self-image concerns in the dictator game, we apply a similar treatment variation to that of Bicchieri et al. (2023). We vary the subjects' ability to manipulate their beliefs by informing them about the course of the experiment once before and once after the belief elicitation task. We hypothesize that, if given the chance, some subjects manipulate their beliefs to behave selfishly while maintaining a positive self-image. For those subjects, the preference for the dictator role increases. Therefore, if belief manipulation is possible, we expect more subjects to strategically choose pessimistic beliefs, behave selfishly, and bid higher as they value the dictator role higher. When belief manipulation is impossible, more subjects maintain optimistic beliefs, behave prosocially, and bid lower as they value the dictator role lower. This leads to our first three hypotheses.

Hypothesis Belief 1. *More subjects choose pessimistic beliefs when belief manipulation is possible.*

Hypothesis Belief 2. *More subjects behave selfishly in the dictator game when belief manipulation is possible.*

Hypothesis Belief 3. *Bids are higher in the second-price auction when belief manipulation is possible.*

In our second approach to control self-image concerns, we apply a similar treatment as the *hidden information* treatment from Dana et al. (2007). We vary the subjects' ability to remain strategically ignorant. A 2x2 matrix design combines the opportunity for strategic ignorance with each information condition. We hypothesize that, if given the chance, some subjects choose strategic ignorance to behave selfishly while maintaining a positive self-image. For those subjects, the preference for the dictator role increases. Therefore, if strategic ignorance is possible, we expect just as many or fewer subjects to maintain pessimistic beliefs than under the respective *open* condition, because strategic ignorance is costless compared to belief manipulation. Ignorant subjects will choose allocation A more often than subjects who reveal matrix 1 and those who have neither of the two strategies in the *open* × *after* treatment. Ignorant subjects value the dictator role more and bid higher than those who reveal the active matrix and those who have neither of the two strategies in the *open* × *after* treatment.

Hypothesis Ignorance 1. *Just as many or fewer subjects choose pessimistic beliefs when strategic ignorance is possible.*

Hypothesis Ignorance 2. *Ignorant subjects more often behave selfishly than subjects who reveal matrix 1 or have neither of the two strategies in the open × after treatment.*

Hypothesis Ignorance 3. *Ignorant subjects bid higher in the second-price auction than subjects who reveal the matrix or have neither of the two strategies in the open × after treatment.*

4.5 Analysis

The experiment was run on oTree via Prolific[6] in June 2023 over several days. Participants were located in the UK and had to fulfill certain criteria [7]. Overall, 332 subjects participated in the experiment. Table A1 in the Appendix provides an overview of how subject characteristics are distributed across the different treatments. Each treatment includes at least 79 and at most 90 observations. The male-to-female ratio is nearly equal across all treatments and, rounded to whole numbers, splits the data evenly. The average age is 38.43 over all treatments and differs slightly between them. Subjects most often reported their income to be in the range of 20.000-29.999 USD, which did not differ between the treatments. Altruism is based on a survey measure by Falk et al. (2018). The measure is calculated as the weighted average of the answers' z-scores to two hypothetical questions on the allocation of money. We find no significant difference between the *open* treatments. Still, one between the *hidden* treatments ($p = 0.069$, t-test) with altruism being about 0.20 points larger in the *hidden* × *after* treatment[8]. In the two *open* treatments, payoffs correspond with matrix 1. In the *hidden* treatments, a random draw decides whether payoffs follow matrix 1 or 2. In the *hidden* × *before* treatment matrix 1 was active in 41.46% of the cases, in the *hidden* × *after* treatment in 54.32% of the cases. The average

[6]Prolific is an online platform for interactive and survey experiments that recruits its subject pool from around the world.

[7]fluent in English, aged between 18 and 99, a Prolific internal approval rate of 95%, previous submissions between 5 and 10.000, only allowed to participate in one treatment.

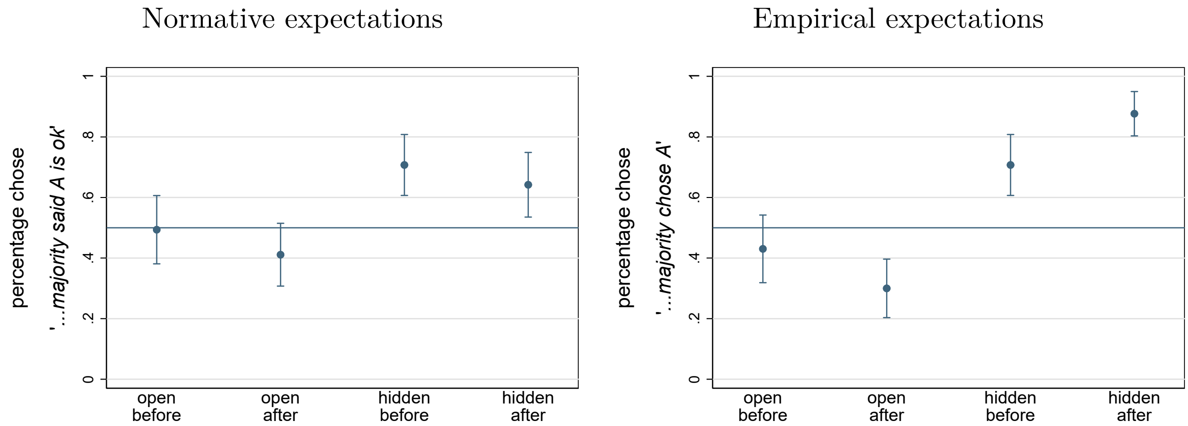
[8]The altruism survey was completed after the experiment. We cannot rule out that subjects engaged in the survey strategically, but lack the data to investigate.

payoff is £2.86, which differs slightly across treatments. An average participation time over all treatments of 14:48 minutes amounts to an average hourly wage of £11.60. Prolific advises an hourly wage of £9, therefore payoffs were above average. Between treatments, we observe a clear difference in the time required to complete the study. Especially, both *hidden* treatments took about 3 minutes longer to complete, which is in line with their increased complexity. Despite their longer duration, average hourly wages remain above £10.

In this section, we present the results of the experiment. We begin by summarizing how normative and empirical expectations changed between the treatments. Then we take a closer look at the dictator decision. Finally, we come to the primary variable of interest - bidding behavior in the auction.

4.5.1 Belief Elicitation Task

In the first step of the analysis, we summarize the subjects' normative and empirical expectations in the belief elicitation task across treatments. Both expectations are dichotomous choice variables. We report frequencies and apply the χ^2-test to asses differences between the treatments.



mean	0.49	0.41	0.71	0.64
freq.	39/79	37/90	58/82	52/81

mean	0.43	0.30	0.71	0.88
freq.	34/79	27/90	58/82	71/81

Figure 4.5: **Normative and empirical expectations across treatments.** This figure presents the mean values and 95% confidence intervals of participants' normative expectations (left panel) and their empirical expectations (right panel) across all four treatments. Normative expectations refer to subjects' expectations about whether the majority in a prior study 'said it is OK to choose A' in the dictator game. Empirical expectations refer to subjects' expectations about whether the majority in a prior study 'chose A' in the dictator game. In the previous study, the majority chose "B" in the *open* treatments and "A" in the *hidden* treatments. The means and frequencies are displayed below the graphs.

Figure 4.5 includes two panels summarizing the mean values and 95% confidence intervals of normative and empirical expectations across the four treatments. The left panel concerns normative expectations and depicts the percentage of subjects who choose "...majority said A is ok" across the four treatments. Normative expectations are more often pessimistic in the *hidden*

treatments. The difference amounts to 21 percentage points (p = 0.006, χ^2-test) under the *before* condition and 23 percentage points (p = 0.003, χ^2-test) under the *after* condition. Within *open* and *hidden* conditions, expectations are less often pessimistic in the *after* treatments. Under the *open* condition, the difference is 8 percentage points (p = 0.282, χ^2-test) and under the *hidden* condition, it is 6 percentage points (p = 0.373, χ^2-test). However, both differences lack statistical significance.
The right panel in figure 4.5 is concerned with empirical expectations and depicts the percentage of subjects who choose "...majority chose A" over the four treatments. Compared with normative expectations, empirical expectations differ significantly across both dimensions. Again, expectations are more often pessimistic in the *hidden* treatments. Under the *before* condition, the difference between *open* and *hidden* treatments is 28 percentage points ($p < 0.001$, χ^2-test), and under the *after* condition, it is 57 percentage points ($p < 0.001$, χ^2-test). Within the *open* and *hidden* conditions, the difference between *before* and *after* treatments is significant as well, but the direction is exactly opposite. Under the *open* condition, the difference is 13 percentage points (p = 0.078, χ^2-test), and under the *hidden* condition, it is 17 percentage points (p = 0.008, χ^2-test). Under the *open* condition, more subjects are pessimistic in the *before* treatment. Notably, under the *hidden* condition, more subjects are pessimistic in the *after* treatment.

Table A2 in the appendix presents the results of a linear regression analysis of normative and empirical expectations. We apply linear probability models with normative and empirical expectations as the respective dependent variables and include various control variables. Because the dependent variable is dichotomous, we also report the results of a logistic regression in table A3 in the Appendix as a robustness check. Regression analysis supports the findings from the non-parametric analysis above.

We summarize this part of the analysis in two findings. First, under the *open* condition, more subjects hold pessimistic expectations in the *before* than in the *after* treatment. Although we find this shift in both expectations, it is only marginally significant in empirical expectations. This is consistent with the results from Bicchieri and Xiao (2009), who identified the importance of empirical expectations in belief manipulation. The regression analysis identifies the exact marginally significant change in empirical expectations. Despite the finding being marginally significant, it supports hypothesis Belief 1 and provides evidence for belief manipulation. Second, in both expectations, we observe a shift towards more pessimistic expectations from the *open* to the *hidden* condition. This finding contradicts hypothesis Ignorance 1. We expect subjects to substitute belief manipulation with costless strategic ignorance. Instead, beliefs have turned even more pessimistic. This change aligns with the prior study's change, which incentivizes the elicitation task in this main study. In the *open* treatment from the previous study, a majority expects the majority to choose the statement in favor of allocation B, and in the *hidden* treatment, in favor of allocation A. Therefore, under each condition, the majority in the main study merely reflects the correct majorities from the prior study. On closer inspection, we notice that introducing strategic ignorance via the *hidden* condition also adds additional variation to the belief elicitation task. Under the *hidden* condition, a random draw chooses between the known matrix 1 and a new matrix 2 with equal probability. Matrix 2 inverts the recipient's payoffs from matrix 1, and allocation A becomes better than B for both roles. Since subjects were asked to make their guesses independent of whether the game was revealed, their assessments are made under uncertainty and include the state of the world in which matrix 2 is active. This is a different question from the one asked under the *open* condition, and therefore, the answers are not comparable. Unfortunately, this design flaw makes it difficult to analyze strategic ignorance in isolation. Subjects are inadvertently but correctly equipped with beliefs advantageous to evade self-image concerns without the need for either strategic ignorance or belief manipulation.

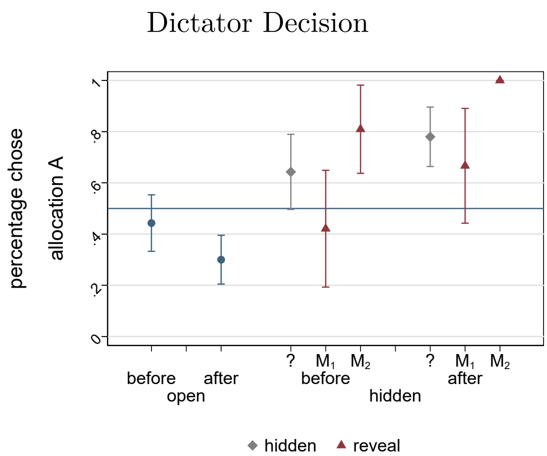
Finding Belief 1. *Under the open condition, empirical expectations are more often pessimistic in the before than in the after treatment.*

Finding Ignorance 1. *Under the hidden condition, expectations are more often pessimistic than in the respective open treatments. This shift is likely caused by a design flaw in the belief elicitation task's questions, which adds additional variation between treatments.*

4.5.2 Dictator Game

In this second step of the analysis, we summarize the subjects' decisions in the dictator game. We gathered the dictator decisions from every subject by applying the strategy method to maximize the number of observations. In the *hidden* treatments, a random draw activates one of the two matrices. Subjects decide whether to reveal the active matrix before making their dictator decision. The dictator and reveal decision variables are both dichotomous. We apply the χ^2-test to assess differences between treatments. We start the analysis by summarizing the reveal decision. Next, we present the dictator decision across the *open* treatments and dependent on the reveal decision in the *hidden* treatments. Finally, we apply simple regression analysis to test the effect of the reveal decision and expectations on the dictator decision across treatments.

The reveal decision precedes the dictator decision and is only made in the *hidden* treatments. In the *before* treatment, 49 percent of subjects reveal the active matrix (40/82). Of those, 47 percent reveal matrix 1 (19/40) and 52 percent reveal matrix 2 (21/40). In the *after* treatment, 38 percent reveal the active matrix (31/81). Of those, 58 percent reveal matrix 1 (18/31) and 42 percent reveal matrix 2 (13/31). The rate of revealing subjects differs between the *hidden* treatments by 10 percentage points (p = 0.176, χ^2-test), but this difference is insignificant.



mean	0.44	0.30	0.64	0.42	0.81	0.78	0.67	1.00
freq.	35/79	27/90	27/42	8/19	17/21	39/50	12/18	13/13

Figure 4.6: **Dictator Decision across Treatments.** This figure presents the mean values and 95% confidence intervals of participants' dictator decision across all treatments. In the *hidden* treatments, the dictator decision is conditioned on whether the active matrix stays hidden (diamond) or is revealed (triangle) and whether matrix 1 (M_1) or matrix 2 (M_2) is active. The means and frequencies are displayed below the graph.

Figure 4.6 presents the mean values and 95% confidence intervals of the subjects' dictator decisions across the four treatments. In the *hidden* treatments, the dictator decision is conditioned on whether the active matrix stays hidden (diamond) or is revealed (triangle), and whether matrix 1 (M_1) or matrix 2 (M_2) is active. Under the *open* condition, the proportion of dictators choosing allocation A increases from *after* to the *before* treatment by 14 percentage points (p = 0.054, χ^2-test). Subjects more frequently choose the selfish option in the *before* treatment. Under the *hidden* condition, we broadly observe an increase in subjects choosing allocation A. Ignorant subjects in both *hidden* treatments choose allocation A more often than the informed subjects in the *open* × *after* treatment, who have neither of the two avoidance strategies available. The difference is 34 percentage points compared to the *before* treatment ($p < 0.001$, χ^2-test) and 48 percentage points compared to the *after* treatment ($p < 0.001$, χ^2-test). Ignorant subjects more often choose allocation A than subjects of the same treatment who reveal matrix 1, but the differences are insignificant. The difference is 22 percentage points in the *before* treatment ($p = 0.105$, χ^2-test) and 11 percentage points in the *after* treatment ($p = 0.341$, χ^2-test). Between the two *hidden* treatments, the difference is 14 percentage points ($p = 0.146$, χ^2-test). Interestingly, we find that subjects who reveal matrix 1 in the *hidden* × *after* treatment more often choose allocation A than the informed subjects in the *open* × *after* treatment. The difference is 37 percentage points ($p = 0.003$, χ^2-test). Both groups are informed about their recipient's payoff consequences, still, we observe a much greater propensity to take advantage of the dictator role under the *hidden* condition. This cannot be explained by strategic ignorance alone. Similar to the expectations data, the dictator data hints at an additional effect introduced by the *hidden* treatment. For completeness, we note that subjects who reveal matrix 2 most often and expectedly choose the efficient allocation A. In the *after* treatment, all 13 subjects choose allocation A. In the *before* treatment, 4/21 subjects choose allocation B.

	open × *before*		*open* × *after*		*hidden* × *before*		*hidden* × *after*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normative	0.238**	0.189	0.0719	-0.000542	0.254*	0.250*	0.0418	0.0939
	(0.107)	(0.122)	(0.0921)	(0.0958)	(0.140)	(0.147)	(0.118)	(0.118)
Empirical	0.319***	0.241*	0.460***	0.430***	0.368**	0.333**	0.344*	0.221
	(0.108)	(0.124)	(0.107)	(0.111)	(0.140)	(0.157)	(0.177)	(0.211)
Reveal					-0.239*	-0.149	-0.119	-0.0353
					(0.132)	(0.141)	(0.125)	(0.130)
Constant	0.188**	0.419	0.132**	0.464***	0.181	0.0423	0.457**	0.500*
	(0.0782)	(0.296)	(0.0529)	(0.164)	(0.138)	(0.332)	(0.172)	(0.263)
Controls	no	yes	no	yes	no	yes	no	yes
Observations	79	71	90	86	61	56	68	61
R^2	0.183	0.238	0.228	0.370	0.226	0.285	0.094	0.215

Table 4.2: **Linear Regression: Expectations and Reveal Decision on Dictator Decision.** This table shows the estimated coefficients from linear regressions of normative expectations, empirical expectations, and the reveal decision on the binary variable dictator decision across the four treatments. The reveal decision only occurs in the *hidden* treatments. The "Dictator" decision takes on the value 1 whenever a subject decides for allocation A. In both *hidden* treatments (5) - (8), we excluded all observations in which matrix 2 was revealed, as this dissolves the social dilemma. Each model includes one pure specification and one with control variables. The following controls were included: Age, Female, Income, Altruism, Totalapprovals, Timetaken. Robust standard errors are shown in parentheses. Some of the controls were not measured for all participants. Differences in sample size between treatments and lack of control variables for some observations explain the difference in sample size. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.2 summarizes the OLS coefficients of two linear probability models for each treatment. We regress normative expectations, empirical expectations, and in the *hidden* treatments, the reveal decision on the binary variable dictator decision across the four treatments. One specification includes a set of control variables, and the other only the primary variables of interest. Adding the control variables results in noticeably higher R^2 values, indicating better model fit across all treatments. We excluded observations in which subjects revealed matrix 2 in the *hidden* treatments, as this solves the social dilemma. In both *open* treatments (see specifications (1)-(4)), the coefficients on empirical expectations suggest a positive effect on the likelihood of a subject choosing allocation A in the dictator game. An additional regression analysis in table A5 in the appendix compares the effect of each expectation on the dictator decision between the *before* and *after* treatments in the *open* and *hidden* conditions. We find no clear indication that the impact of expectations on the dictator decision between the *open* treatments differs (see the interaction terms in specifications (1) and (2)). The data suggest that empirical expectations have an indistinguishable positive effect on the dictator decision in both *open* treatments. In the *hidden* treatments, the effect of expectations is slightly weaker. We find positive coefficients throughout both *hidden* treatments. The significance is weaker, and in specification (8) it is lost. The reveal decision has a negative coefficient and is only marginally significant in the *before* treatment in the model without controls (5). The additional regression analysis in table A5 in the appendix does not indicate that either the impact of expectations or that of the reveal decision differs between the treatments. In an exploratory attempt, we interact the explanatory variables with different control variables, as their inclusion explained a noticeable increase in the data's variation. Table A6 in the appendix presents linear regression models interacting the explanatory variables with the demographic variable *age*. Again, R^2 increases noticeably, and the interaction uncovers how the effect of empirical expectation, in all but the *open* × *after* treatment, significantly increases with age. We specifically display the interaction with age, as age can serve as a proxy for life experience, which has a profound effect on human behavior. Our results from this exploratory attempt further substantiate the importance of empirical expectation on the dictator decision.

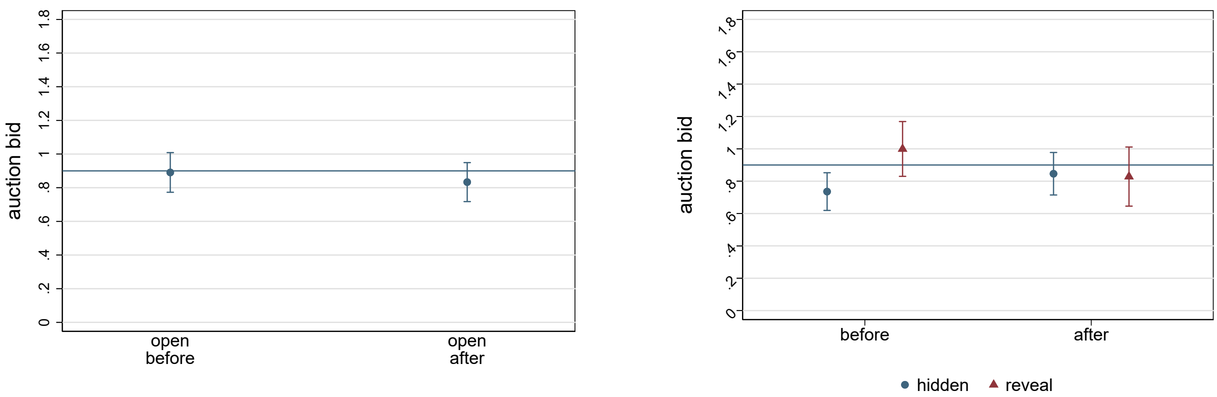
We summarize the results of this subsection in two key findings.
First, under the *open* condition, we find an increase in selfish choices in the *before* compared to the *after* treatment. This finding aligns with hypothesis Belief 2. Regression analysis identifies empirical expectations as a strong predictor of selfish behavior in the dictator decision. Combined with finding 1, the increase of pessimistic empirical expectations in the *before* treatment, the data provides strong evidence for belief manipulation. Second, about half of the subjects choose ignorance in each *hidden* treatment. Ignorant subjects significantly more often choose allocation A than informed subjects in the *open* × *after* treatment who have neither of the two avoidance strategies available. Ignorant subjects more often choose allocation A than subjects who reveal matrix 1 in the respective *hidden* treatment, but the differences are insignificant. Ignorant subjects more often choose allocation A in the *after* than in the *before* treatment. Regression analysis unveils a positive effect of empirical expectations and no effect of the reveal decision on the dictator decision. Therefore, the shift towards allocation A cannot distinctly be attributed to strategic ignorance. Also, subjects who reveal matrix 1 behave more selfishly than informed subjects in the *open* × *after* treatment. Their behavior cannot be explained by strategic ignorance alone and therefore another force must be at play in the *hidden* treatments. Considering the accumulation of pessimistic expectations under the *hidden* condition and the positive effect of expectations on the dictator decision suggests that the shift in behavior towards allocation A is driven by the changes in expectations and may be attributable to the design flaw.

Finding Belief 2. *Under the open condition, subjects behave more selfishly in the before treatment. Pessimistic empirical expectations are a strong predictor of selfishness in both open treatments and appear more often in the before treatment. Combined, this presents strong evidence for belief manipulation.*

Finding Ignorance 2. *Under the hidden condition, about half of the subjects choose ignorance* *in both treatments. Ignorant subjects of both hidden treatments more often choose allocation A than informed subjects in the open × after treatment. Pessimistic empirical expectations have a positive effect on the reveal decision, and the reveal decision does not affect the dictator decision. Therefore, the shift towards allocation A cannot distinctly be attributed to strategic ignorance.*

4.5.3 Auction Bid

Finally, we attend to the subjects' bidding behavior in the second-price auction. We interpret variations in behavior as an indication of a change in the preference for the dictator role. The variable auction bid can take on discrete values between 0 and 1.8 with 0.01 steps. It can be treated as a continuous variable in approximation. We present the p-values from the two-sample t-test and compare the entire distributions using the Kolmogorov-Smirnov test. Further, we apply the Mann-Whitney U test whenever the normality assumption fails[9] or when the sample size is too low to justify a t-test. First, we summarize the second-price auction bids across treatments. Next, we deepen our analysis by conditioning on the dictator decision.



mean	0.89	0.83
obsv.	79	90

mean	0.73	1.00	0.85	0.83
freq.	42	40	50	31

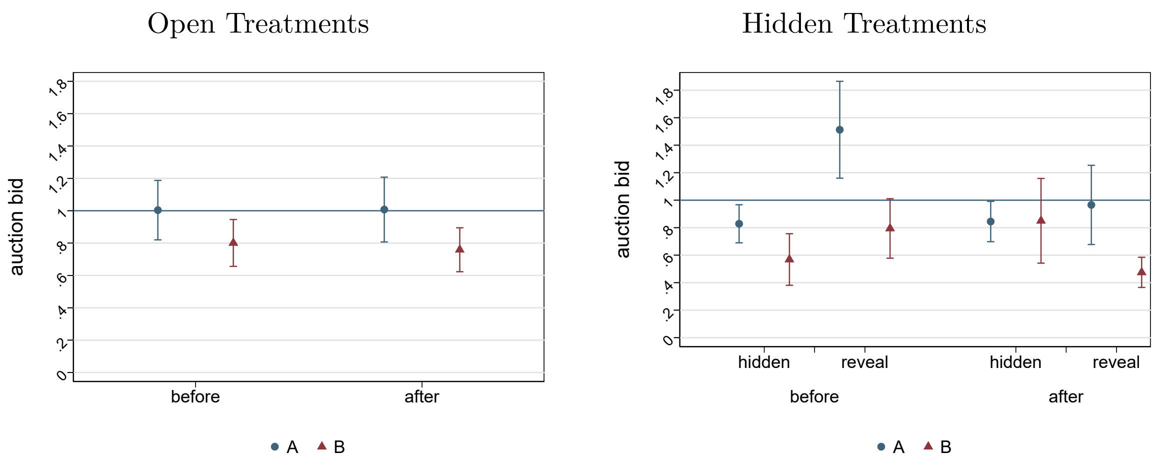
Figure 4.7: **Auction Bid across Treatments.**This figure presents the mean values and 95% confidence intervals of participants' second-price auction bids in the *open* treatments (left panel) and in the *hidden* treatments divided according to the reveal decision (right panel). The right panel distinguishes between auction bids from ignorant subjects (dots) and those from subjects who reveal the active matrix (triangles). Subjects do not know which matrix is active at the time of the auction. Therefore, it is irrelevant which matrix is active. Bids can take discrete values from 0.00 to 1.80, in steps of 0.01. A blue line at 0.90 illustrates bidding half of what is possible. The means and frequencies are displayed below the graphs.

Figure 4.7 includes two panels summarizing the mean values and 95% confidence intervals of auction bids across the four treatments. The left panel presents auction bids in the *open* treatments. Average auction bids are slightly higher in the *before* treatment, but the difference of £0.06 ($p = 0.492$, t-test) is insignificant. The distributions do not differ between the treatments ($p = 0.780$, Kolmogorov-Smirnov test).
The right panel presents auction bids for the two *hidden* treatments divided up according to the reveal decision. By game-theoretic reasoning, we assume that subjects decide on their bid with a complete plan of action in mind, especially concerning the reveal decision. At the time of the auction, subjects do not know which matrix is active. Therefore, we pool the bids of revealing subjects regardless of the active matrix. Ignorant subjects' average auction bids neither exceed those of revealing subjects under the respective *hidden* treatment, nor those of informed subjects in the *open* × *after* treatment. On the contrary, ignorant subjects in the *before* treatment bid,

[9]We test for normality by the Shapiro-Wilk test.

on average, £0.26 less than revealing subjects ($p = 0.014$, t-test), and their distributions differ significantly ($p = 0.049$, Kolmogorov-Smirnov test). Compared to informed subjects in the *open* × *after* treatment, ignorant subjects bid on average insignificantly £0.10 less ($p = 0.2436$, t-test) and their distributions do not differ ($p = 0.162$, Kolmogorov-Smirnov test). In the *after* treatment, the average bids and distributions of ignorant subjects do not significantly differ from those of either group of informed subjects.

Based on this first look at the data, we must reject both of our hypotheses on auction bids. We find no convincing evidence that belief manipulation increases the first-mover preference under the *open* condition. Under the *hidden* condition, strategic ignorance has no effect in the *after* treatment and coincides with even lower auction bids in the *before* treatment. We further investigate by including subjects' behavior in the dictator game. This allows us to account for differences in motivations.



mean	1.00	0.80	1.01	0.76
obsv.	35	44	27	63

before

mean	0.83	0.57	1.51	0.80
obsv.	27	15	8	11

after

mean	0.84	0.85	0.96	0.47
freq.	39	11	12	6

Figure 4.8: **Auction Bid across Treatments by Dictator and Reveal Decision.** This figure presents the mean values and 95% confidence intervals of participants' second-price auction bids for the *open* treatments depending on their dictator decision (left panel) and the *hidden* treatments depending on their dictator and reveal decision (right panel). In the left panel, we separate means by subjects choosing allocation A (dots) and allocation B (triangles). In the right panel, we separate means on two levels. First, whether subjects remain ignorant or reveal the active matrix. Second, whether they choose allocation A (dots) or allocation B (triangles). The reveal decision takes place before the dictator decision. The dictator decision, therefore, depends on which matrix is active. We are only interested in subjects' bidding behavior when the social dilemma is (potentially) active. Therefore, we exclude subjects who reveal matrix 2. Bids can take discrete values from 0.00 to 1.80, in steps of 0.01. A blue line at 0.90 illustrates bidding half of what is possible. The means and numbers of observations are displayed below the graphs.

Figure 4.8 includes two panels summarizing the mean values and 95% confidence intervals of auction bids across the four treatments depending on the dictator and the reveal decision. The left panel presents auction bids across the *open* treatments, depending on the dictator decision, separating subjects who choose allocation A (dots) and allocation B (triangles). In both treatments, selfish subjects bid marginally significantly higher on average. The difference is £0.20 in the *before* treatment ($p = 0.089$, t-test) and £0.25 in the *after* treatment ($p = 0.046$, Mann-Whitney U test). Their distributions are only suggestively different (*before*: $p = 0.104$; *after*: $p = 0.114$, Kolmogorov-Smirnov test). Between the treatments, neither the average bids ($p = 0.492$, t-test) nor the distributions ($p = 0.780$, Kolmogorov-Smirnov test) differ significantly for selfish subjects. The right panel presents auction bids across the *hidden* treatments dependent on the dictator and the reveal decision. In each treatment, subjects are divided according to their reveal decision and whether they choose allocation A (dot) or allocation B (triangle). The reveal decision takes place before the dictator decision. Therefore, the choice in the dictator decision depends on which matrix is active. We are only interested in subjects' bidding behavior when the social dilemma is (potentially) active. Therefore, we exclude subjects who reveal matrix 2. The combination of conditioning and exclusion results in very few observations (n<30). We apply the Mann-Whitney U test to compare distributions. In the *before* treatment, ignorant subjects bid, on average, £0.26 more when they choose allocation A, and the distributions differ marginally (p = 0.068, Mann-Whitney U test). Compared to revealing subjects who choose allocation A, ignorant subjects who choose allocation A bid on average £0.68 less, and their distributions differ significantly (p = 0.001, Mann-Whitney). Compared to informed subjects who choose allocation A in the *open* × *after* treatment, ignorant subjects who choose allocation A bid on average £0.17 less, but their distributions do not differ (p = 0.194, Mann-Whitney U test). In the *after* treatment, ignorant subjects show no difference in bidding behavior concerning their dictator decision (p = 0.903, Mann-Whitney U test). Compared to revealing subjects who choose allocation A, ignorant subjects who choose allocation A bid on average £0.12 less, but the distributions do not differ significantly (p = 0.352, Mann-Whitney U test). Compared to informed subjects who choose allocation A in the *open* × *after* treatment, ignorant subjects who choose allocation A bid on average £0.16 less, but their distributions do not differ (p = 0.203, Mann-Whitney U test).

We summarize this part of the analysis in two findings. First, under the *open* condition, average auction bids are slightly, but insignificantly higher in the *before* treatment. Average auction bids are higher among subjects who behave selfishly in both treatments. Those selfish subjects bid, on average, the same, and their bid distributions do not differ across treatments.
Second, under the *hidden* condition, the results for ignorant subjects contradict hypothesis Ignorance 3. In the *after* treatment, average bids of ignorant subjects do not differ from revealing subjects and the informed subjects in the *open* × *after* treatment. In the *before* treatment, ignorant subjects bid less than the revealing subjects and insignificantly less to the informed subjects in the *open* × *after* treatment. In the *after* treatment, the behavior of ignorant subjects does not differ. We further divide the subjects by their dictator decision to control for intentions. This results in very low sample sizes. Ignorant subjects who choose allocation A bid on average more in the *before* and the same in the *after* treatment. Surprisingly, ignorant subjects bid differently according to their intention only in the *before* treatment. Comparing the bid behavior of ignorant subjects who choose allocation A to revealing subjects and informed subjects in the *open* × *after* treatment who also choose allocation A, we do not find the hypothesized increase. Instead, we find mild evidence for a decrease.

Finding Belief 3. *Under the open condition, average auction bids are slightly, but insignificantly higher in the before treatment. Selfish subjects bid on average higher in both treatments. Average bids and distributions of selfish subjects are the same in both treatments.*

Finding Ignorance 3. *In the hidden treatment, ignorant subjects bid, on average, less than revealing subjects and insignificantly less than informed subjects. In the after treatment, ignorant subjects bid on average the same as revealing and informed subjects. In the before treatment, ignorant subjects who choose allocation A bid higher than those who choose allocation B. In the after treatment, ignorant subjects bid about the same amount, independent of their dictator decision. In the before treatment, ignorant subjects who choose allocation A bid less than revealing and insignificantly less than informed subjects who choose allocation A. In the after treatment, ignorant subjects who choose allocation A bid similarly to revealing and informed subjects.*

4.6 Discussion and Conclusion

In this article, we investigate the connection between self-image concerns and the first-mover preference in bargaining. We design an online experiment based on the first-mover preference elicitation strategy by Litsios and Schories (2024). They combine a dictator game with a prefixed second-price auction, selling the dictator role. Changes in bidding behavior between treatments reveal changes in the dictator role preference. Further, we adjust their setup to study the role of self-image concerns. We applied two treatment variations that enable subjects to avoid concerns about their self-image and engage in selfish behavior. First, as in Bicchieri et al. (2023), we vary subjects' ability to strategically manipulate their beliefs by altering the game's information structure. Second, as in Dana et al. (2007), we vary the subjects' ability to remain strategically ignorant of their recipient's payoff consequences.

The analysis leads to two main results. First, we find mild evidence that belief manipulation increases the first-mover preference, and therefore, self-image concerns play a role in who makes the first offer. Second, our attempt to isolate the effect of strategic ignorance was unsuccessful because a design flaw introduced additional variation between the treatments. Strategic ignorance coincides with a drastic shift towards allocation A, while bidding behavior reveals no change in the first-mover preference. In the following, we summarize the findings and outline the chain of reasoning that leads to our main results.

Under the *open* condition, we observe an accumulation of pessimistic empirical expectations in the *before* treatment alongside an increase in selfish behavior. Regression analysis further provides evidence that empirical expectations are correlated with selfishness in both treatments. These findings suggest a strong influence of expectations and imply that some subjects respond to the *before* treatment with belief manipulation. Average auction bids are slightly, but insignificantly, higher in the *before* treatment. Selfish subjects bid significantly higher in both treatments. Their bidding behavior is indistinguishable between the treatments. Although the *before* treatment contains more selfish subjects, the distributions do not differ. This suggests that subjects who respond to the *before* treatment express no systematically different valuation of the dictator role than selfish subjects in the *after* treatment. We presume that the lack of significance between the *open* treatments is due to a limited number of self-image-concerned subjects who respond to the incentives in our parametrization and to the dichotomous character of the elicitation task and the dictator game. Nevertheless, the increase in selfish behavior and the unchanged distribution of selfish subjects' bids support our claim: belief manipulation serves to evade self-image concerns, enables selfish behavior, and increases first-mover preferences; therefore, self-image concerns play a role in who makes the first offer.

Against our hypothesis, we observe that normative and empirical expectations are more frequently pessimistic in both *hidden* treatments. We attribute this drastic shift in expectations primarily to the design flaw in the belief elicitation task. Under the *hidden* condition, we asked subjects about their beliefs independent of their reveal decision or the active matrix. The question includes uncertainty and thereby differs from the one asked under the *open* condition. The design flaw introduces additional variation and inadvertently endows subjects with advantageous beliefs that promote selfish behavior. In both *hidden* treatments, about half the subjects choose ignorance. Ignorant subjects more often choose allocation A than informed subjects in the *open* × *after* treatment. There is no significant difference between ignorant subjects and subjects who reveal matrix 1 in either *hidden* treatment. Allocation A does not necessarily lead to an unequal outcome. But refraining from costless revelation of the active matrix and ignorantly choosing allocation A most closely resembles selfishness. Further, regression analysis reveals a positive effect of empirical expectations and no effect of the reveal decision on the dictator decision. Therefore, the shift towards allocation A cannot distinctly be attributed to strategic ignorance. If strategic ignorance alone caused the change in dictator behavior, then ignorant subjects should choose allocation A more often than subjects who reveal matrix 1, which is not the case. Also, subjects who reveal matrix 1 in the *hidden* × *after* treatment more often choose allocation A than informed subjects in the *open* × *after* treatment. These two groups have the same information and face the same social dilemma. The groups differ in the greater prevalence of pessimistic beliefs in the *hidden* treatment. This suggests that the pessimistic expectations affect the shift towards allocation A in the *hidden* treatments. We cannot distinctly attribute the behavior change to either expectations or ignorance, but we acknowledge that their mutual appearance coincides with increased selfishness. Bidding behavior of ignorant subjects contradicts our hypothesis Ignorance 3. Ignorant subjects bid the same or possibly less than revealing and informed subjects. This decrease is only significant compared to revealing subjects in the *before* treatment, but averages hint at a broader tendency. One possible explanation for the decline in bidding behavior lies in an appreciation of the recipient role under the *hidden* condition. The expected payoff of the recipient increases due to matrix 2, which most likely results in recipients receiving the high payoff. As the expected payoff difference between the two roles diminishes, the dictator role becomes relatively less attractive, leading to a decrease in average bids.

The current study provides some insights into our research question, but we encounter several challenges and shortcomings that shall be addressed in the following. First, we observe belief manipulation, but the data only unveils a mild effect on the first-mover preference. The parametrization of our experiment, along with the dichotomous character of the elicitation task and the dictator game, provides a specific frame for observing a behavior change. There are only two possible allocations in the dictator game. One is completely equal, and the other is at the high end of inequality, with the recipient receiving only 15 percent of the surplus[10]. The elicitation task focuses on the majority, which demands strong convictions and rules out more subtle adjustments. If there is an individual constraint on the extent of belief manipulation, then our setup can only induce an observable behavioral change for a limited set of characteristics. As an extreme example of a constraint on belief manipulation, there will be very few of us who could genuinely convince ourselves into believing that the majority would take the entire surplus. Besides larger changes in the elicitation task and dictator game, a less unequal allocation A would be the simplest way to address the issue. Second, the design flaw in the elicitation task prevents us from isolating the effect of strategic ignorance. One way to address this issue would be to extend the elicitation task under the *hidden* condition by asking questions that depend on each matrix. Subjects do not need to aggregate both contingencies in one belief at this early stage. It would allow us to directly compare the beliefs between the *open* and *hidden* conditions and test for belief manipulation under strategic ignorance. Subjects would likely be more focused on the ignorance decision itself. Third, we must question whether strategic ignorance, as we have used it, is an appropriate approach to measure changes in the first-mover preference. Ignorant subjects know that by a 50 percent chance, matrix 2 is active, and the recipient role becomes more attractive. A possible increase in first-mover preference driven by diverted self-image concerns might be offset by a decrease in the preference due to a higher expected payoff as a recipient. These opposing forces complicate the comparison of auction bids between the *open* and *hidden* conditions. However, comparing auction bids between ignorant and revealing subjects could still be insightful. Fourth, the entire experiment is lengthy and complex. Some subjects found the instructions lengthy and complex, while others found them intriguing. We addressed comprehension by using control questions and by offering the correct solution when incorrect answers were given. We allowed subjects to click through the instructions with no risk of adverse consequences, and we did not collect data on the number of incorrect answers. In retrospect, our approach failed to collect valuable data on subjects' comprehension and is much too lenient for online experiments. One advantage of online experiments is that, compared to lab experiments, a larger pool of subjects is available to participate at a given time. Threatening exclusion if a

[10]Engel (2011) find that dictators share, on average, 28.35% of the surplus

certain number of control questions are answered incorrectly would have been a sensible way to incentivize attentive participation and exclude unsuitable individuals.

The study enriches the discussion on the first-offer dilemma by adding a behavioral economics perspective. We show how self-image concerns can affect who makes the first move in negotiations. Despite the effect on auction behavior being small, the data show that the ability to avoid self-image concerns positively affects the prospective earnings an individual can justify to themselves. In the absence of an avoidance strategy, self-image concerns effectively structure the first-mover preference by constraining the justifiable earnings. These concerns are relevant even under the strictest allocation of bargaining power, such as in a dictator game. Even though a dictator can freely decide without regard for strategic considerations, it is their own identity that limits their behavior. Even though the design flaw in the *hidden* treatment hinders our understanding of the role of strategic ignorance, it shows how a combination of the proper set of beliefs and a self-chosen lack of knowledge can almost entirely undermine prosocial behavior. This finding casts substantial doubt on the existence of a context-independent preference for inequality. Belief manipulation alone resulted in less selfish behavior than its combination with strategic ignorance. We presume the existence of individual constraints that restrict the extent to which beliefs can be manipulated. Given the importance of beliefs to self-image concerns, it would be worthwhile to investigate the constraints on belief manipulation further. The results of this study suggest considering the framing of the situation and the individual characteristics of the negotiation parties when evaluating who will make the first move. One should ask questions like: How important is identity? What are the identities? What are common beliefs? How susceptible to manipulation are those beliefs? The results also highlight the valuable signaling effect of the first offer. The informational advantage of the second mover over the first mover's preference provides a rationale for not making the first move. This poses the question: When is the benefit of making the first move greater than the advantage of receiving information as the second-mover? This aspect can play a considerable role when assessing a potential employer(/e) or business partner. Leaving the first offer to the other party might come at the expense of anchoring in a conventional understanding of negotiations. On the other hand, it might save you a lot of time and hassle to know early on in the relationship how your opposite prefers to split profits. In conclusion, our findings illustrate that the first-mover preference is neither purely a matter of financial strategic reasoning nor of innate preferences alone. As we remove the financial strategic considerations, we observe that conveying the preferred self-concept remains a crucial factor beyond a simple fairness preference. Self-image concerns can determine not only what a dictator offers, but also affect who wants to make this offer in the first place. Introducing this psychological dimension into our understanding of negotiation dynamics can make a difference in leading successful negotiations and maybe even help us determine who to negotiate with in the first place. Or as a famous saying goes, "When you talk, you repeat what you already know; when you listen, you often learn something."

Bibliography

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, 115(3):715–753.

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.

Babcock, L., Loewenstein, G., Issacharoff, S., and Camerer, C. (1995). Biased judgments of fairness in bargaining. *The American Economic Review*, 85(5):1337–1343.

Bem, D. J. (1972). Self-perception theory. *Advances in experimental social psychology*, 6.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4):848–886.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American economic review*, 96(5):1652–1678.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–64.

Bicchieri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press.

Bicchieri, C., Dimant, E., and Sonderegger, S. (2023). 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.

Bodner, R. and Prelec, D. (2003). Self-signaling and diagnostic utility in everyday decision making. *The psychology of economic decisions*, 1(105):26.

Bohnet, I. and Frey, B. S. (1999). Social distance and other-regarding behavior in dictator games: Comment. *American Economic Review*, 89(1):335–339.

Bolton, G. E. and Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193.

Brandts, J. and Charness, G. (2011). The strategy versus the direct-response method: a first survey of experimental comparisons. *Experimental Economics*, 14(3):375–398.

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton university press.

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.

Cotter, M. J. and Henley Jr, J. A. (2008). First-offer disadvantage in zero-sum game negotiation outcomes. *Journal of Business-to-Business Marketing*, 15(1):25–44.

Dana, J., Cain, D. M., and Dawes, R. M. (2006). What you don't know won't hurt me: Costly (but quiet) exit in dictator games. *Organizational Behavior and human decision Processes*, 100(2):193–201.

Dana, J., Weber, R. A., and Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. *Economic Theory*, 33(1):67–80.

Dell, D. and Boswell, J. (2009). *Never make the first offer:(Except when you should) wisdom from a master dealmaker*. Penguin.

Di Tella, R., Perez-Truglia, R., Babino, A., and Sigman, M. (2015). Conveniently upset: avoiding altruism by distorting beliefs about others' altruism. *American Economic Review*, 105(11):3416–42.

Dufwenberg, M. and Muren, A. (2006). Generosity, anonymity, gender. *Journal of Economic Behavior & Organization*, 61(1):42–49.

Engel, C. (2011). Dictator games: A meta study. *Experimental economics*, 14:583–610.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D., and Sunde, U. (2018). Global Evidence on Economic Preferences. *The Quarterly Journal of Economics*, 133(4):1645–1692.

Fehr, E. and Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868.

Galinsky, A. D. and Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology*, 81(4):657–669.

Gangadharan, L., Grossman, P. J., and Xue, N. (2024). Belief elicitation under competing motivations: Does it matter how you ask? *European Economic Review*, page 104830.

Gino, F., Norton, M. I., and Weber, R. A. (2016). Motivated bayesians: Feeling moral while acting egoistically. *Journal of Economic Perspectives*, 30(3):189–212.

Grossman, Z. and Van Der Weele, J. J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association*, 15(1):173–217.

Gunia, B. C., Swaab, R. I., Sivanathan, N., and Galinsky, A. D. (2013). The remarkable robustness of the first-offer effect: Across culture, power, and issues. *Personality and Social Psychology Bulletin*, 39(12):1547–1558.

Güth, W. and Tietz, R. (1986). Auctioning ultimatum bargaining position. *Current Issues in West German Decision Research, Lang, Frankfurt, Germany*, pages 60–73.

Hoffman, E., McCabe, K., and Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American economic review*, 86(3):653–660.

Kahneman, D., Knetsch, J. L., Thaler, R., et al. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American economic review*, 76(4):728–741.

Karni, E. (2009). A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606.

Krupka, E. L. and Weber, R. A. (2013). Identifying social norms using coordination games: Why does dictator game sharing vary? *Journal of the European Economic Association*, 11(3):495–524.

Litsios, C. and Schories, F. (2024). Making the first offer in negotiation – explaining role preferences by social image concerns in the dictator game.

Loschelder, D. D., Swaab, R. I., Trötschel, R., and Galinsky, A. D. (2014). The first-mover disadvantage: The folly of revealing compatible preferences. *Psychological science*, 25(4):954–962.

Loschelder, D. D., Trötschel, R., Swaab, R. I., Friese, M., and Galinsky, A. D. (2016). The information-anchoring model of first offers: When moving first helps versus hurts negotiators. *Journal of Applied Psychology*, 101(7):995–1012.

McCormack, M. H. (2022). *What they still don't teach you at Harvard Business School*. Profile Books.

Norgaard, K. M. (2006). "we don't really want to know" environmental justice and socially organized denial of global warming in norway. *Organization & Environment*, 19(3):347–370.

Ochs, J. and Roth, A. E. (1989). An experimental study of sequential bargaining. *The American Economic Review*, 79(3):355–384.

Orr, D. and Guthrie, C. (2005). Anchoring, information, expertise, and negotiation: New insights from meta-analysis. *Ohio State Journal on Dispute Resolution*, 21:597–628.

Osório, A. (2020). On the first-offer dilemma in bargaining and negotiations. *Theory and Decision*, 89(2):179–202.

Rubinstein, A. (1982). Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society*, 50(1):97–109.

Shachat, J. and Swarthout, J. T. (2013). Auctioning the right to play ultimatum games and the impact on equilibrium selection. *Games*, 4(4):738–753.

Appendix

A Additional Tables

Demographics

	N	Female	Age	Income*	Altruism	Matrix 1	Payoff	Time
private-before	79	48%	36.72	25	-0.06	100%	2.79	12:36
private-after	90	49%	37.82	25	0.00	100%	2.96	13:34
hidden-before	82	50%	40.11	25	-0.07	41.46%	2.84	16:40
hidden-after	81	49%	39.09	25	0.13	54.32%	2.83	16:24

Table A1: **Summary of participant demographics.** This table summarizes participant demographics across treatments. Treatments were randomly allocated. The self-reported income is elicited in brackets, and we report the respective midpoints, so that income bracket 25, for example, means that a participant's yearly income is between USD 10.000 and USD 19.999. Altruism is measured using a survey instrument by Falk et al. (2018). The measure is calculated as the weighted average of the answers' z-scores to two hypothetical questions on the allocation of money. In the two *open* treatments, matrix 1 is always active. In the *hidden* treatments, a random draw decides whether matrix 1 or 2 is active. The payoff is reported in £, and the average time to finish the experiment is given in (min:sec) format. Note that *N* represents the total number of observations in our sample, while for calculating the summary statistics, a few participants' demographics are missing in the data set.

	Normative			Empirical		
	(1)	(2)	(3)	(4)	(5)	(6)
Hidden	0.223***	0.214***	0.205**	0.431***	0.277***	0.263***
	(0.0532)	(0.0759)	(0.0811)	(0.0489)	(0.0755)	(0.0806)
After	-0.0741	-0.0826	-0.0961	0.0170	-0.130*	-0.147*
	(0.0532)	(0.0770)	(0.0811)	(0.0493)	(0.0742)	(0.0781)
Hidden × After		0.0172	0.0446		0.300***	0.314***
		(0.107)	(0.112)		(0.0970)	(0.103)
Constant	0.489***	0.494***	0.621***	0.352***	0.430***	0.432***
	(0.0479)	(0.0566)	(0.115)	(0.0468)	(0.0560)	(0.106)
Controls	no	no	yes	no	no	yes
Observations	332	332	307	332	332	307
R^2	0.057	0.057	0.074	0.189	0.212	0.220

Table A2: **Linear Regression: Normative Expectations and Empirical Expectations.** This table shows the estimated coefficients from linear regressions of normative expectations (1, 2, 3) and empirical expectations (4, 5, 6) on binary variables indicating the particular treatment. Normative expectations refer to subjects' expectations about whether the majority of subjects in a prior study 'said it is OK to choose A' in the dictator game. Empirical expectations refers to subjects' expectation about whether the majority of subjects in a prior study 'chose A' in the dictator game. "hidden" takes the value 1 for the set of *hidden* treatments (T3 and T4) and zero otherwise; "After" takes the value 1 for the set of *after* treatments (T2 and T4) and zero otherwise. In (3) and (6) the following controls were included: Age, Female, Income, Fair, Totalapprovals, Timetaken. Robust standard errors are shown in parentheses. Some of the controls were not measured for all participants. This explains the difference in observations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	Normative			Empirical		
	(1)	(2)	(3)	(4)	(5)	(6)
Hidden	0.926***	0.908***	0.890**	1.909***	1.163***	1.125***
	(0.229)	(0.331)	(0.355)	(0.250)	(0.333)	(0.357)
After	-0.318	-0.334	-0.397	0.0859	-0.567*	-0.655*
	(0.228)	(0.311)	(0.332)	(0.248)	(0.324)	(0.346)
Hidden × After		0.0356	0.157		1.645***	1.689***
		(0.458)	(0.483)		(0.528)	(0.551)
Constant	-0.0339	-0.0253	0.523	-0.617***	-0.280	-0.275
	(0.196)	(0.225)	(0.492)	(0.216)	(0.228)	(0.516)
Controls	no	no	yes	no	no	yes
Observations	332	332	307	332	332	307
Pseudo R^2	0.0421	0.0421	0.0557	0.1443	0.1669	0.1735

Table A3: **Logistic Regression: Normative expectations and empirical expectations.** This table shows the estimated coefficients from logistic regressions of normative expectations (1, 2, 3) and empirical expectations (4, 5, 6) on binary variables indicating the particular treatment. Normative expectations refer to subjects' expectations about whether the majority of subjects in a prior study 'said it is OK to choose A' in the dictator game. Empirical expectations refers to subjects' expectation about whether the majority of subjects in a prior study 'chose A' in the dictator game. "hidden" takes the value 1 for the set of *hidden* treatments (T3 and T4) and zero otherwise; "After" takes the value 1 for the set of *after* treatments (T2 and T4) and zero otherwise. In (3) and (6) the following controls were included: Age, Female, Income, Fair, Totalapprovals, Timetaken. Robust standard errors are shown in parentheses. Some of the controls were not measured for all participants. This explains the difference in observations. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	open × before		open × after		hidden × before		hidden × after	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normative	1.108**	0.972	0.429	-0.0187	1.190*	1.496*	0.231	0.631
	(0.507)	(0.610)	(0.533)	(0.723)	(0.665)	(0.821)	(0.650)	(0.712)
Empirical	1.426***	1.185**	2.153***	2.601***	1.690**	1.736**	1.579**	0.948
	(0.503)	(0.591)	(0.535)	(0.747)	(0.721)	(0.784)	(0.768)	(1.033)
Reveal					-1.125*	-0.706	-0.654	-0.0769
					(0.637)	(0.732)	(0.645)	(0.751)
Constant	-1.431***	-0.329	-1.839***	0.674	-1.479**	-2.862	-0.164	-0.0645
	(0.456)	(1.388)	(0.405)	(1.240)	(0.754)	(1.939)	(0.717)	(1.489)
Controls	no	yes	no	yes	no	yes	no	yes
Observations	79	71	90	86	61	56	68	61
Pseudo R^2	0.1407	0.1955	0.1809	0.3538	0.1747	0.2498	0.0756	0.2062

Table A4: **Logistic Regression: Expectations and Reveal Decision on Dictator Decision.** This table shows the estimated coefficients from logistic regressions of normative expectations, empirical expectations, and the reveal decision on the binary variable dictator decision across the four treatments. The reveal decision only occurs in the *hidden* treatments. The "Dictator" decision takes on the value one whenever a subject decides for allocation A. In both *hidden* treatments (5) - (8), we excluded all observations in which matrix 2 was revealed, as this dissolves the social dilemma. Each model includes one pure specification and one with control variables. The following controls were included: Age, Female, Income, Altruism, Totalapprovals, Timetaken. Robust standard errors are shown in parentheses. Some of the controls were not measured for all participants. Differences in sample sizes across treatments and the lack of control variables for some observations explain the difference in sample sizes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	open condition		*hidden* condition	
	(1)	(2)	(3)	(4)
Before	0.0560	0.0176	-0.128	-0.165
	(0.0944)	(0.111)	(0.213)	(0.216)
Empirical	0.460***	0.446***	0.311*	0.180
	(0.108)	(0.113)	(0.173)	(0.169)
Empirical × Before	-0.141	-0.137	-0.0338	0.0557
	(0.153)	(0.166)	(0.216)	(0.210)
Normative	0.0719	0.0201	0.0463	0.105
	(0.0922)	(0.0956)	(0.0995)	(0.0983)
Normative × Before	0.166	0.182	0.0622	-0.00325
	(0.141)	(0.154)	(0.164)	(0.160)
Reveal			0.0152	0.0697
			(0.143)	(0.143)
Reveal × Before			-0.000404	-0.000971
			(0.222)	(0.228)
Constant	0.132**	0.445***	0.482***	0.451**
	(0.0529)	(0.147)	(0.168)	(0.213)
Controls	no	yes	yes	yes
Observations	71	86	56	61
R^2	0.222	0.255	0.113	0.212

Table A5: **Linear Regression: Expectations, Reveal Decision and Interaction Terms on Dictator Decision.** This table shows the estimated coefficients from linear regressions of the before treatment, normative expectations, empirical expectations, the reveal decision, and their interaction terms with the before treatment on the binary variable dictator decision under the *open* and *hidden* condition. The Before variable is a dummy taking on the value 1 for the *before* treatment. Normative refers to the subjects' normative expectations. Empirical refers to the subjects' empirical expectations. The reveal decision only occurs in the *hidden* treatments. Subjects decide whether they reveal the active payoff matrix or remain ignorant. A random draw determines whether matrix 1 (social dilemma) or matrix 2 (aligned incentives) is active. The dictator decision takes the value one whenever a subject chooses allocation A. Under the *hidden* condition (3) and (4), we excluded all observations where matrix 2 was active because it dissolves the social dilemma. Each model includes one pure specification and one with control variables. The following controls were included: Age, Female, Income, Altruism, Totalapprovals, Timetaken. Robust standard errors are shown in parentheses. Some of the controls were not measured for all participants. Differences in sample sizes across treatments and the lack of control variables for some observations explain the difference in sample sizes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	open × *before*		*open* × *after*		*hidden* × *before*		*hidden* × *after*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Normative	0.189	0.279	-0.000542	0.593*	0.250*	1.012*	0.0939	0.0864
	(0.122)	(0.399)	(0.0958)	(0.301)	(0.147)	(0.529)	(0.118)	(0.394)
Normative × Age		-0.00225		-0.0152**		-0.0176		-0.000355
		(0.00968)		(0.00642)		(0.0125)		(0.00879)
Empirical	0.241*	-0.476	0.430***	-0.0912	0.333**	-0.533	0.221	-0.830
	(0.124)	(0.373)	(0.111)	(0.357)	(0.157)	(0.449)	(0.211)	(0.568)
Empirical × Age		0.0199**		0.0134		0.0214**		0.0289**
		(0.00913)		(0.00827)		(0.0104)		(0.0122)
Reveal					-0.149	-0.914*	-0.0353	-0.219
					(0.141)	(0.535)	(0.130)	(0.488)
Reveal × Age						0.0183		0.00664
						(0.0125)		(0.0117)
Age	-0.00485	-0.0166*	-0.00672**	-0.00345	0.0000568	-0.00439	0.00371	-0.0213**
	(0.00518)	(0.00861)	(0.00313)	(0.00357)	(0.00616)	(0.0136)	(0.00453)	(0.0105)
Constant	0.419	0.850**	0.464***	0.364**	0.0423	0.213	0.500*	1.442***
	(0.296)	(0.398)	(0.164)	(0.175)	(0.332)	(0.551)	(0.263)	(0.494)
Controls	yes	yes	yes	yes	yes	yes	yes	yes
Observations	71	71	86	86	56	56	61	61
R^2	0.238	0.300	0.370	0.421	0.285	0.392	0.215	0.295

Table A6: **Linear Regression: Expectations and Reveal Decision in Interaction with Age on Dictator Decision.** This table shows the estimated coefficients from linear regressions of normative expectations, empirical expectations, and the reveal decision in interaction with age on the binary variable dictator decision across the four treatments. The reveal decision only occurs in the *hidden* treatments. The "Dictator" decision takes on the value one whenever a subject decides for allocation A. In both *hidden* treatments (5) - (8), we excluded all observations in which matrix 2 was revealed, as this dissolves the social dilemma. Each model includes one specification with the control variables and one that interacts the three explanatory variables with the control variable Age. The following controls were included: Age, Female, Income, Altruism, Totalapprovals, Timetaken. Robust standard errors are shown in parentheses. Some of the controls were not measured for all participants. Differences in sample sizes across treatments and the lack of control variables for some observations explain the differences in sample sizes. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B Experimental Instructions

B.1 General Instructions

Welcome Page
In this experiment, we explore human decision-making. You will be asked to complete computer tasks illustrating rather abstract situations. You can earn money, depending on your own as well as other participants' behavior. Be informed that we do NOT use deception as it is sometimes done in other studies. We will also ask you to provide demographic information. We will not ask for your name or any information that will make you identifiable. Overall, this study will take between 10 to 20 minutes depending on the assigned treatment. With your participation in this study, you will receive £1.8. Additionally, you may receive a monetary bonus depending on your and others' decisions and the assigned treatment. The exact amount depends on your results in the experiment. The risks from participating are no greater than those encountered in everyday life decision-making. Your participation in this study is completely voluntary, and you may refuse to participate or withdraw from the study without penalty or loss of benefits to which you may otherwise be entitled. Compensation will be awarded upon completion of the entire study. We are very glad that you chose to participate in this study and hope that your participation will be an interesting experience for you. In the end, we will provide you with some further details on the objective of this study. If you would like to know even more, please feel free to contact us using the following email address: christos.litsios@uni-hamburg.de Enjoy!

Consent Form
Please read the following consent form before continuing: I consent to participate in this research study. I am free to withdraw at any time without giving a reason, knowing that any payments only become effective if I complete the study. I understand that all data will be kept confidential by the researchers. Individual names and other personally identifiable information are not available to the researchers and will not be asked at any time. No personally identifiable information will be stored with or linked to data from the study. I consent to the publication of study results as long as the information is anonymous so that no identification of participants can be made. I hereby confirm that I am at least 18 years old. I have read and understood the explanations above and I voluntarily consent to participate in this study.

General Information
The study consists of three parts:

- Part A: introduction to an earlier study
- Part B: incentivized tasks,
- Part C: a questionnaire.

Part A: introduction to an earlier study In Part A you will be presented an earlier study carried out on prolific.
In Part B you will be asked to complete incentivized tasks concerned with the matter of the earlier study.
In Part C you will be asked to complete a questionnaire on demographic variables as well as some preferences.
After the instructions in Part B, you will have to answer a set of comprehension questions. You will earn £1.8 for participating, and you can earn additional money (bonus payment) in this study based on your and the other participants' decisions in Part B. You will be paid the total

amount through Prolific within 48 hours, only if you have entirely completed the study. If you are participating on your tablet or phone, we ask you to flip your device such that all pages are shown in landscape mode. When you fully understood the instructions, please click the "Next"-button!

Introduction to Earlier Study
Please read the instructions of the earlier study carefully! General Information

- The study consists of three parts:
  - Part 1: guessing task,
  - Part 2: decision-making task,
  - Part 3: a questionnaire.
- In the *guessing* task (Part 1), you will learn about the decision-making task from Part 2 and then be asked to guess other participants' behavior as well as their opinion on different behaviors.
- In the *decision-making* task (Part 2), you are matched with another participant. Both of you make a decision about the allocation of a sum of money.

B.2 Open Treatments Instructions

Instructions Decision Task Earlier Study
Set-Up

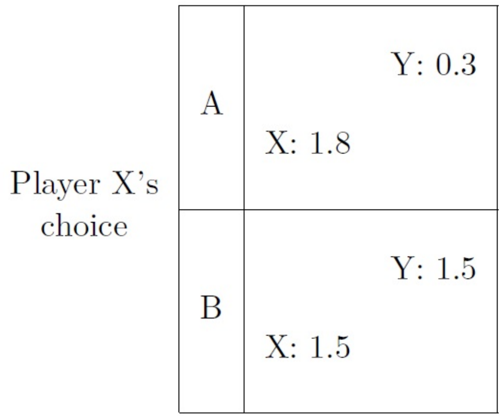
- Two participants from Prolific are matched randomly.
- Both remain completely anonymous to each other.

Timing

- Both make a decision conditional on becoming Player X in the decision-making task described below.
- After the decision, a random draw (50:50 chance) assigns the role of Player X and Player Y.
- The decision of the participant who has been assigned the role of Player X will be executed.

Decision-Making Task

- There are two roles, Player X and Player Y.



- Player X can choose between two allocations, namely A and B.
  - A: Player X gets £1.8 and Player Y gets £0.3
  - B: Player X gets £1.5 and Player Y gets £1.5
- Player Y receives the respective amount in correspondence with Player X's decision.

Instructions Guessing Task Earlier Study

- The guessing task consists of two assessments about the decision-making task in Part 2.
  - One assessment about the opinion of the majority of participants in this study.
  - One assessment about the behavior of the majority of participants in this study.
- Each assessment comes in the form of a question about which of two mutually exclusive statements is true. Like:
  - In this study, most people *say it is OK to choose* A / B.
  - In this study, most people *choose* A / B.
- You decide which statements you think are true.
- For each correct assessment you receive a bonus payment of £0.1.
- "Correct" means:
  - ...in the opinion assessment, you guess the statement that was most frequently chosen in the guessing task.
  - ...in the behavior assessment, you guess the behavior that was most frequently chosen in the decision-making task.

Open × After Treatment

Instructions Guessing Task in Current Study
Please read these instructions carefully! Your task in this current study:

- Now, after you have learned about the guessing and decision-making tasks from the earlier study (Part A), you will engage in a guessing task yourself.

Guessing Task in Current Study

- The guessing task is similar to the one from the earlier study. It only differs in that now you are guessing opinion and behavior from the earlier study and not from the current study.
- Therefore, you make two assessments, one on the guessing task and one on the decision-making task from the earlier study (Part A).
- Each assessment comes in the form of a question about which of two mutually exclusive statements is true. Like:
  - In the earlier study, most people *said it is OK to choose* A / B — opinion assessment on the guessing task.
  - In the earlier study, most people *chose* A / B — behavior assessment on the decision-making task.
- You decide which statements you think are true.
- For each correct assessment you receive a bonus payment of £0.1.
- "Correct" means:
  - ...in the opinion assessment, you guess the statement that was most frequently chosen in the guessing task of the earlier study.
  - ...in the behavior assessment, you guess the behavior that was most frequently chosen in the decision-making task of the earlier study.

Instructions on Decision Task in Current Study
Instructions on Task in Current Study Please read these instructions carefully! Another task in this current study: Now after you have finished the *guessing task*, you will participate in a *decision-making task* like the one from the earlier study (Part 2). This *decision-making task* differs from the one in the earlier study in one crucial aspect: You can now increase the probability of becoming Player X by using your initial endowment and making the highest bid in an *auction* preceding the decision-making task. Auction and Decision-Making Task in Current Study Set-Up

- Two participants from the current study are matched randomly.
- Both remain completely anonymous to each other.

Timing

- First, the two participants compete in an auction. The winner of the auction increases their chance to become Player X up to 90% probability. The loser becomes Player X with 10% probability.
- Second, both make a decision *conditional* on becoming Player X in the decision-making task described below. *At the time of the decision you do not know your role yet!*
- After the decision, a random draw assigns the role of Player X and Player Y. The auction winner has a 9:1 chance of becoming Player X.
- The decision of the participant who has been assigned the role of Player X will be executed.

Auction

- The participants can use their initial endowment of £1.80.
- They can bid any amount between £0.00 and £1.80 (using two decimals, including £0.00 and £1.80).
- The bidder with the highest bid wins the auction, but only pays the second-highest bid.
- One can bid their true valuation for the role of Player X without running the risk of paying more than is needed to win the auction.
- The bidder with the second-highest bid loses the auction but keeps their entire endowment of £1.80 independent of the amount.
- In the highly unlikely case of a draw (both participants bid the same amount), a 50:50 chance decides who wins the auction. The winner pays their bid, the loser pays nothing.
- The auction winner becomes Player X with 90% probability and Player Y with 10% probability. Winner keeps = £1.80 - bid_{2nd}.
- The auction loser becomes Player X with 10% probability and Player Y with 90% probability. Loser keeps = £1.80.
- At the time you make your choice in the decision-making task, you neither know whether you won the auction, what you paid, nor which role you have. You will only learn this at the very end of the study.

Decision-Making Task

- There are two roles, Player X and Player Y.

(Here would appear the payoff table image.)

- Player X can choose between two allocations, A and B.
  - A: Player X gets £1.80 and Player Y gets £0.30.
  - B: Player X gets £1.50 and Player Y gets £1.50.
- Player Y receives the respective amount based on Player X's decision.

When you fully understood the instructions, please click the "Next" button as soon as it appears.

Open × Before Treatment

Instructions Guessing Task in Current Study
Please read these instructions carefully! Your task in this current study:

- Now, after you have learned about the *guessing* and *decision-making* tasks from the earlier study (Part A), you will engage in a guessing task yourself.

Guessing Task in Current Study

- The guessing task is similar to the one from the earlier study. It only differs in that now you are guessing opinion and behavior from participants of the earlier study and not from the current study.
- There are two assessments, one on the guessing task and one on the decision-making task from the earlier study (Part A).
- Each assessment comes in the form of a question about which of two mutually exclusive statements is true. Like:
  - In the earlier study, most people *said it is OK to choose* A / B — opinion assessment on the guessing task.
  - In the earlier study, most people *chose* A / B — behavior assessment on the decision-making task.
- You decide which statements you think are true.
- For each correct assessment you receive a bonus payment of £0.10.
- "Correct" means:
  - ...in the opinion assessment, you guess the statement that was most frequently chosen in the guessing task of the earlier study.
  - ...in the behavior assessment, you guess the behavior that was most frequently chosen in the decision-making task of the earlier study.

When you fully understood the instructions, please click the "Next" button as soon as it appears.

Instructions Decision Task Study in Current Study
Please read these instructions carefully! Another task in this current study:

- After you have finished the *guessing task*, you will participate in a *decision-making task* like the one from the earlier study (Part 2).
- This *decision-making task* differs from the one in the earlier study in one crucial aspect:
  - You can now increase the probability of becoming Player X by using your initial endowment and making the highest bid in an *auction* preceding the decision-making task.

Auction and Decision-Making Task in Current Study Set-Up

- Two participants from the current study are matched randomly.
- Both remain completely anonymous to each other.

Timing

- First, the two participants compete in an auction. The winner of the auction increases their chance to become Player X up to 90% probability. The loser becomes Player X with 10% probability.

- Second, both make a decision *conditional* on becoming Player X in the decision-making task described below. *At the time of the decision you do not know your role yet!*
- After the decision, a random draw assigns the role of Player X and Player Y. The auction winner has a 9:1 chance of becoming Player X.
- The decision of the participant who becomes Player X will be executed.

Auction

- The participants can use their initial endowment of £1.80.
- They can bid any amount between £0.00 and £1.80 (using two decimals, including both £0.00 and £1.80).
- The bidder with the highest bid wins the auction, but only pays the second-highest bid.
- One can bid their true valuation for the role of Player X without running the risk of paying more than is necessary to win the auction.
- The bidder with the second-highest bid loses the auction but keeps their entire endowment of £1.80.
- In the highly unlikely case of a draw (both participants bid the same amount), a 50:50 chance decides who wins the auction. The winner pays their bid, the loser pays nothing.
- The auction winner becomes Player X with 90% probability and Player Y with 10% probability. Winner keeps = £1.80 - bid_{2nd}.
- The auction loser becomes Player X with 10% probability and Player Y with 90% probability. Loser keeps = £1.80.
- At the time you make your choice in the decision-making task, you neither know whether you won the auction, what you paid, nor which role you have. You will only learn this at the very end of the study.

Decision-Making Task

- There are two roles, Player X and Player Y.
- Player X can choose between two allocations:
  - A: Player X gets £1.80 and Player Y gets £0.30.
  - B: Player X gets £1.50 and Player Y gets £1.50.
- Player Y receives the respective amount based on Player X's decision.

When you fully understood the instructions, please click the "Next" button as soon as it appears.

B.3 Hidden Treatments Instructions

Instructions Decision Task Earlier Study
Introduction to Earlier Study Please read the instructions of the earlier study carefully! Instructions Decision-Making Task **Set-Up**

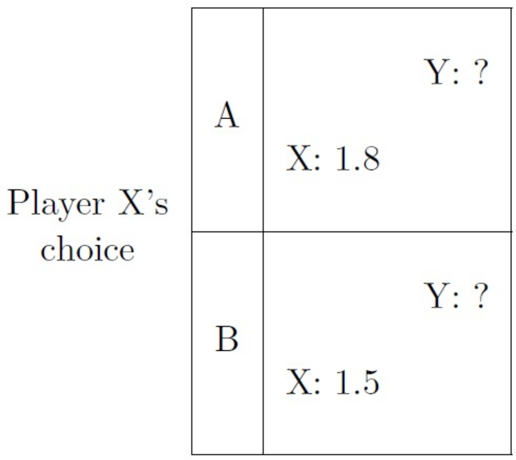
- Two participants from Prolific are matched randomly.
- Both remain completely anonymous to each other.

Timing

- Both make a decision *conditional* on becoming Player X in the decision-making task described below.
- After the decision, a random draw (50:50 chance) assigns the role of Player X and Player Y.
- The decision of the participant who has been assigned the role of Player X will be executed.

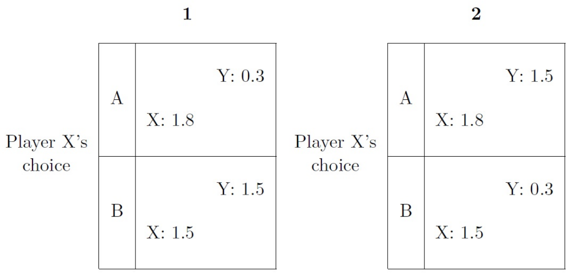
Decision-Making Task

- There are two roles, Player X and Player Y.



- In the above matrix Player Y's payoff is hidden.
- Player X is presented the payoff matrix and a button that reveals the game.
- Before Player X makes an allocation choice, Player X can choose whether to reveal the game and uncover Player Y's payoff *at no cost*, or stay ignorant about Player Y's payoff.
- Independent of Player X's choice, chance decides which of the two matrices, either matrix 1 or matrix 2, is active. The program chooses between the two options with equal probability (50:50).

- If Player X chooses to reveal the game, Player X is presented the corresponding active matrix.
- If Player X chooses not to reveal the game, Player X makes the allocation choice and stays forever ignorant about the payoff consequences for Player Y.



Therefore one of the following three scenarios may materialize:

- If Player X chooses not to reveal the game, Player X can directly choose between:
  - A: Player X gets £1.80 and stays ignorant about Player Y's payoff, which is £0.30 with 50% probability and £1.50 with 50% probability.
  - B: Player X gets £1.50 and stays ignorant about Player Y's payoff, which is £1.50 with 50% probability and £0.30 with 50% probability.
- If Player X chooses to reveal the game and the program selects matrix 1 to be active, Player X chooses between:
  - A: Player X gets £1.80 and Player Y gets £0.30.
  - B: Player X gets £1.50 and Player Y gets £1.50.
- If Player X chooses to reveal the game and the program selects matrix 2 to be active, Player X chooses between:
  - A: Player X gets £1.80 and Player Y gets £1.50.
  - B: Player X gets £1.50 and Player Y gets £0.30.

- Player Y receives the respective amount corresponding to Player X's decision.

Instructions Guessing Task Earlier Study
Please read the instructions of the earlier study carefully! Instructions Guessing Task

- The guessing task consists of two assessments about the decision-making task in Part 2.
  - One assessment about the *opinion* of the majority of participants in this study.
  - One assessment about the *behavior* of the majority of participants in this study.
- Each assessment comes in the form of a question about which of two mutually exclusive statements is true:
  - In this study, most people *say it is OK to choose* A / B (opinion assessment).

  - In this study, most people *choose* A / B (behavior assessment).

- You decide which statements you think are true.
- For each correct assessment you receive a bonus payment of £0.10.
- "Correct" means:
  - ...in the *opinion* assessment, you guess the statement that was most frequently chosen in the guessing task.
  - ...in the *behavior* assessment, you guess the behavior that was most frequently chosen in the decision-making task.

Hidden × After Treatment

Instructions Guessing Task Current Study
Instructions on Task in Current Study Please read these instructions carefully! Your task in this current study:

- Now, after you have learned about the guessing and decision-making tasks from the earlier study (Part 1 and 2), you will engage in a guessing task yourself.

- The guessing task is similar to the one from the earlier study. It only differs in that now you are guessing opinion and behavior from participants of the earlier study and not from the current study.
- There are two assessments, one on the guessing task and one on the decision-making task from the earlier study (Part A).
- Each assessment comes in the form of a question about which of two mutually exclusive statements is true. For example:
  - In the earlier study, most people *said it is OK to choose* A / B — opinion assessment on the guessing task.
  - In the earlier study, most people *chose* A / B — behavior assessment on the decision-making task.
- You decide which statements you think are true.
- For each correct assessment you receive a bonus payment of £0.10.
- "Correct" means:
  - ...in the opinion assessment, you guess the statement that was most frequently chosen in the guessing task of the earlier study.
  - ...in the behavior assessment, you guess the behavior that was most frequently chosen in the decision-making task of the earlier study.

When you fully understood the instructions, please click the "Next" button as soon as it appears.

Instructions Decision Task Current Study
Instructions on Task in Current Study Please read these instructions carefully! Another task in this current study:

- After you have finished the *guessing task*, you will participate in a *decision-making task* like the one from the earlier study (Part 2).
- This decision-making task differs from the earlier study in one crucial aspect:

  You can now increase the probability of becoming Player X by using your initial endowment and making the highest bid in an *auction* preceding the decision-making task.

Auction and Decision-Making Task in Current Study Set-Up

- Two participants from Prolific are matched randomly.
- Both remain completely anonymous to each other.

Timing

- First, the two participants compete in an auction. The winner of the auction increases their chance to become Player X up to 90%. The loser only becomes Player X with 10%.
- Second, both make a decision *conditional* on becoming Player X in the decision-making task described below. *At the time of the decision you do not know your role yet.*
- After the decision, a random draw assigns the roles of Player X and Player Y. The auction winner has a 9:1 chance of becoming Player X.
- The decision of the participant assigned the role of Player X will be executed.

Auction

- Participants can use their initial endowment of £1.80.
- They can bid any amount between £0.00 and £1.80 (two decimals, inclusive).
- The participant with the highest bid wins the auction, but only pays the second-highest bid.
  - One can bid their true valuation for the role of Player X without risking to pay more than necessary to win.
- The bidder with the second-highest bid loses the auction but keeps their entire endowment of £1.80 regardless of their bid.
- In the highly unlikely case of a draw (both participants bid the same amount), a 50:50 chance decides the winner. The winner pays their bid, the loser pays nothing.
- The auction winner becomes Player X with 90% probability and Player Y with 10%. Winner keeps = £1.80 - bid_{2nd}.
- The auction loser becomes Player X with 10% probability and Player Y with 90%. Loser keeps = £1.80.
- At the time you make your choice in the decision-making task, you do not know whether you won the auction, what you paid, or which role you have. You will only learn this at the very end of the study.

Decision-Making Task

- There are two roles: Player X and Player Y.

Player X's choice	A	X: 1.8, Y: ?
	B	X: 1.5, Y: ?

- In the initial matrix, Player Y's payoff is hidden.
- Player X sees the payoff matrix and a button that reveals the game.
- Before making an allocation choice, Player X can choose whether to reveal the game and learn Player Y's payoff at no cost, or remain ignorant.
- Regardless of Player X's reveal choice, the computer randomly (50:50) selects one of two matrices (matrix 1 or matrix 2).
- If Player X reveals, the active matrix is shown.
- If Player X does not reveal, Player X remains ignorant forever and still must choose between A and B.

1

Player X's choice	A	X: 1.8, Y: 0.3
	B	X: 1.5, Y: 1.5

2

Player X's choice	A	X: 1.8, Y: 1.5
	B	X: 1.5, Y: 0.3

Therefore, one of the following scenarios can occur:

- If Player X chooses *not* to reveal the game:
  - A: Player X receives £1.80 and is ignorant of Player Y's payoff, which is £0.30 with 50% and £1.50 with 50%.

  - B: Player X receives £1.50 and is ignorant of Player Y's payoff, which is £1.50 with 50% and £0.30 with 50%.
- If Player X chooses to reveal the game and the active matrix is matrix 1:
  - A: Player X receives £1.80, Player Y receives £0.30.
  - B: Player X receives £1.50, Player Y receives £1.50.
- If Player X chooses to reveal the game and the active matrix is matrix 2:
  - A: Player X receives £1.80, Player Y receives £1.50.
  - B: Player X receives £1.50, Player Y receives £0.30.

- Player Y receives the amount resulting from Player X's decision.
- Player X is completely free to decide. The consequences have no implications beyond this task.

When you fully understand the instructions, please click the "Next" button as soon as it appears.

Hidden × Before Treatment

Instructions Guessing Task Current Study
Please read these instructions carefully! Your task in this current study:

- Now, after you have learned about the *guessing* and *decision-making* tasks from the earlier study (Part 1 & 2), you will engage in a **guessing task** yourself.

- The guessing task is similar to the one from the earlier study. It only differs in that now you are guessing opinion and behavior from participants of the earlier study and not from the current study.
- There are two assessments, one on the *guessing task* and one on the *decision-making task* from the earlier study (Part A).
- Each assessment comes in the form of a question about which of two mutually exclusive statements is true. Like:
  - In the earlier study, most people *said it is OK to choose* A / B — opinion assessment on the guessing task.
  - In the earlier study, most people *chose* A / B — behavior assessment on the decision-making task.
- You decide which statements you think are true.
- For each correct assessment you receive a bonus payment of **£0.10**.
- "Correct" means:
  - ...in the opinion assessment, you guess the statement that was most frequently chosen in the guessing task of the earlier study.
  - ...in the behavior assessment, you guess the behavior that was most frequently chosen in the decision-making task of the earlier study.

When you fully understood the instructions, please click the "Next" button as soon as it appears!

Instructions Decision Task Current Study
Please read these instructions carefully! Another task in this current study:

- After you have finished the *guessing task*, you will participate in a *decision-making task* like the one from the earlier study (Part 2).
- This decision-making task differs from the earlier study in one crucial aspect:

  You can now increase the probability of becoming **Player X** by using your **initial endowment** and making the highest bid in an *auction* preceding the decision-making task.

Auction and Decision-Making Task in Current Study Set-Up

- Two participants from Prolific are matched randomly.
- Both remain completely anonymous to each other.

Timing

- First, the two participants compete in an **auction**. The winner increases their chance to become Player X up to 90%. The loser becomes Player X with only 10%.
- Second, both make a decision *conditional* on becoming Player X in the decision-making task described below. *At the time of the decision, you do not know your role yet.*
- After the decision, a random draw assigns the roles of Player X and Player Y. The auction winner has a 9:1 chance of becoming Player X.
- The decision of the participant who becomes Player X will be executed.

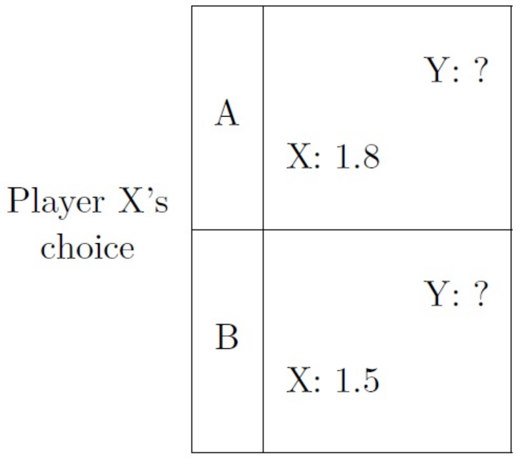
Auction

- Participants can use their initial endowment of £1.80.
- They can bid any amount between £0.00 and £1.80 (two decimals, inclusive).
- The highest bidder wins the auction but only pays the second-highest bid.
  - One can bid their true valuation for the role of Player X without risking to pay more than necessary.
- The second-highest bidder loses the auction but keeps their entire endowment of £1.80.
- In the unlikely case of a draw (both bid the same amount), a 50:50 chance determines the winner. The winner pays their own bid; the loser pays nothing.
- The winner becomes Player X with 90% probability and Player Y with 10%. Winner keeps $= £1.80 - \text{bid}_{\text{2nd}}$.
- The loser becomes Player X with 10% probability and Player Y with 90%. Loser keeps = £1.80.

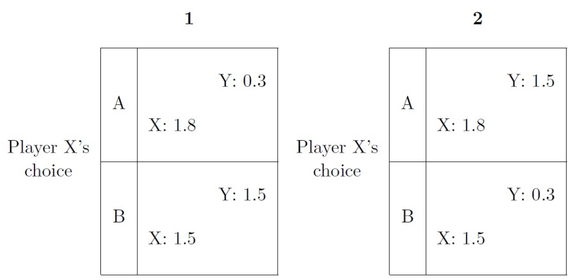
- At the time of making your decision in the decision-making task, you do not know whether you won the auction, what you paid, or which role you will have. You learn this only at the end of the study.

Decision-Making Task

- There are two roles: Player X and Player Y.



- In the initial matrix, Player Y's payoff is hidden.
- Player X sees the payoff matrix and a button that reveals the game.
- Before choosing, Player X can reveal the game at no cost or stay ignorant.
- Regardless of the choice, the computer randomly (50:50) selects either matrix 1 or matrix 2.
- If Player X reveals, the active matrix is shown.
- If Player X does not reveal, they stay ignorant and must still choose between A and B.



Possible scenarios

- If Player X does *not* reveal the game:

  - A: Player X receives £1.80 and stays ignorant about Player Y's payoff, which is £0.30 with 50% and £1.50 with 50%.
  - B: Player X receives £1.50 and stays ignorant about Player Y's payoff, which is £1.50 with 50% and £0.30 with 50%.

- If Player X reveals the game and matrix 1 is active:
  - A: Player X receives £1.80, Player Y receives £0.30.
  - B: Player X receives £1.50, Player Y receives £1.50.
- If Player X reveals the game and matrix 2 is active:
  - A: Player X receives £1.80, Player Y receives £1.50.
  - B: Player X receives £1.50, Player Y receives £0.30.

- Player Y receives the amount resulting from Player X's choice.

When you fully understood the instructions, please click the "Next" button as soon as it appears.

Anhang der Dissertation

Liste der aus dieser Dissertation hervorgegangenen Veröffentlichungen

Chapter 2:

Status: publiziert.

Huber, C., Litsios, C., Nieper, A., & Promann, T. (2023). On social norms and observability in (dis) honest behavior. Journal of Economic Behavior & Organization, 212, 1086-1099.

Chapter 3:

Status: Erste Begutachtungsrunde (Experimental Economics)

Chapter 4:

Status: nicht publiziert.

Abstract

This dissertation contains three essays on behavioral economics. All three studies apply the experimental method to examine the role of a psychological motive in a standard economic context. The common theme to all three papers is image concerns and social norms. The chapters differ with respect to economic context and the kind of image concerns.

Chapter 2 focuses on the impact of social norms and social image concerns on (dis)honesty decisions. We experimentally examine their role by manipulating the observability of actions and nudging different norms. We propose that image concerns can only change behavior in the same direction as the relevant social norm. If there exists a social norm for dishonesty, transparency can lead to adverse behavior through image concerns, just as a social norm for honesty would encourage the opposite. We apply a die-rolling task with a factorial treatment design in an online experiment. Our data show that social-norm nudges can effectively mitigate dishonest behavior. There is significantly less dishonesty when participants are presented with an honest trio than with a dishonest one. Our rather anonymous online experiment fails to identify any effect of observability on the (dis)honesty decision.

Chapter 3 investigates the effect of social image concerns on the first-mover preference in negotiations. We ask whether people are less willing to take the potentially advantageous first-mover position when an audience observes their decision. We abstract from complexity by reducing our research to a simplified environment and examine how social image concerns affect the role preferences in a dictator game experiment. To measure the first-mover preference, a second-price auction sells a probability-advantaged lottery ticket for the dictator role between two competing subjects. Our data show that subjects whose choices are being observed by others offer more in the dictator game. Correspondingly, auction bids are lower for those subjects. The main driving factors are one's own fairness concerns and awareness of one's image concerns.

Chapter 4 investigates the effect of self-image concerns on the first-mover preference in negotiations and thereby also offers a behavioral economics perspective on the first-offer dilemma. We use the novel preference-elicitation mechanism first introduced by Litsios and Schories (2021) in an online experiment to understand the impact of self-image concerns on the first-mover preference. We manipulate self-image concerns by controlling participants' options for belief manipulation and strategic ignorance. Our data give rather weak or no support for our hypothesis about the importance of self-image concerns for the first-mover preference. On the level of the dictator game, however, our two manipulation strategies clearly show an impact.

Zusammenfassung

Diese Dissertation enthält drei Aufsätze zum Thema Verhaltensökonomie. Alle drei Studien wenden die experimentelle Methode an, um die Rolle psychologischer Motive in einem standardmäßigen wirtschaftlichen Kontext zu untersuchen. Das gemeinsame Thema aller drei Arbeiten sind Imagebedenken und soziale Normen. Die Kapitel unterscheiden sich hinsichtlich des wirtschaftlichen Kontexts und der Art der Imagebedenken.

Kapitel 2 konzentriert sich auf den Einfluss sozialer Normen und sozialer Imagebedenken auf (un)ehrliche Entscheidungen. Wir untersuchen ihre Rolle experimentell, indem wir die Beobachtbarkeit von Handlungen manipulieren und verschiedene Normen anstoßen. Wir schlagen vor, dass Imagebedenken das Verhalten nur in dieselbe Richtung wie die relevante soziale Norm verändern können. Wenn es eine soziale Norm für Unehrlichkeit gibt, kann Transparenz durch Imagebedenken zu unerwünschtem Verhalten führen, genauso wie eine soziale Norm für Ehrlichkeit das Gegenteil fördern würde. Wir wenden eine Würfelaufgabe mit einem faktoriellen Behandlungsdesign in einem Online-Experiment an. Unsere Daten zeigen, dass soziale Norm-Nudges unehrliches Verhalten wirksam mindern können. Es gibt deutlich weniger Unehrlichkeit, wenn den Teilnehmenden ein ehrliches Trio präsentiert wird, als wenn ein unehrliches präsentiert wird. Unser eher anonymes Online-Experiment kann keinen Einfluss der Beobachtbarkeit auf die Entscheidung für (Un-)Ehrlichkeit feststellen.

Kapitel 3 untersucht den Einfluss sozialer Imagebedenken auf die Präferenz für die Erstbieter-Rolle in Verhandlungen. Wir fragen, ob Menschen weniger bereit sind, die potenziell vorteilhafte Erstbieter-Rolle einzunehmen, wenn ihre Entscheidung von einem Publikum beobachtet wird. Wir abstrahieren von der Komplexität, indem wir unsere Forschung auf eine vereinfachte Umgebung reduzieren, und untersuchen, wie soziale Imagebedenken die Rollenpräferenzen in einem Diktatorspiel-Experiment beeinflussen. Um die Präferenz für die Erstbieter-Rolle zu messen, wird in einer Zweitpreisauktion ein Lotterielos mit Wahrscheinlichkeitsvorteil für die Diktatorrolle zwischen zwei konkurrierenden Probanden verkauft. Unsere Daten zeigen, dass Probanden, deren Entscheidungen von anderen beobachtet werden, im Diktatorspiel mehr anbieten. Entsprechend sind die Auktionsgebote bei diesen Probanden niedriger. Die wichtigsten treibenden Faktoren sind die eigenen Fairnessbedenken und das Bewusstsein für das eigene Image.

Kapitel 4 untersucht den Einfluss von Selbstbildbedenken auf die Präferenz für die Erstbieter-Rolle in Verhandlungen und bietet damit auch eine verhaltensökonomische Perspektive auf das First-Offer-Dilemma. Wir verwenden den neuartigen Präferenzermittlungsmechanismus, der erstmals von Litsios und Schories (2021) in einem Online-Experiment vorgestellt wurde, um den Einfluss von Selbstbildbedenken auf die Präferenz für die Erstbieter-Rolle zu verstehen. Wir manipulieren Selbstbildbedenken, indem wir die Optionen der Teilnehmenden zur Erwartungs-Manipulation und strategischer Ignoranz steuern. Unsere Daten stützen unsere Hypothese über die Bedeutung von Selbstbildbedenken auf die Präferenz für die Erstbieter-Rolle nur schwach oder gar nicht. Auf der Ebene des Diktatorspiels zeigen unsere beiden Manipulationsstrategien jedoch eindeutig Wirkung.

Erklärung

Hiermit erkläre ich, Christos Litsios, dass ich keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die Arbeit wurde nicht schon einmal in einem früheren Promotionsverfahren angenommen oder als ungenügend beurteilt.

Ort/Datum | Unterschrift Doktorand

Eidesstattliche Versicherung

Ich, Christos Litsios, versichere an Eides statt, dass ich die Dissertation mit dem Titel:

„Behavioral Economic Essays on Image Concerns and Social Norms in Moral Decision-Making and Negotiations“

selbst und bei einer Zusammenarbeit mit anderen Wissenschaftlerinnen oder Wissenschaftlern gemäß den beigefügten Darlegungen nach § 6 Abs. 3 der Promotionsordnung der Fakultät für Wirtschafts- und Sozialwissenschaften vom 18. Januar 2017 verfasst habe. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt.

Ort/Datum | Unterschrift Doktorand

Eigenständigkeitserklärung zur Nutzung von gKI (generative künstliche Intelligenz)

Ich versichere, dass ich KI-Tools lediglich als Hilfsmittel genutzt habe und ich bin mir bewusst, dass die Nutzung maschinell generierter Texte keine Garantie für die Qualität von Inhalten und Text gewährleistet.
Ich verantworte die Übernahme jeglicher von mir verwendeter maschinell generierter Textanpassungen vollumfänglich selbst und dokumentiere im Folgenden, für welche Aufgaben ich KI-Tools genutzt habe.
Ich versichere, dass ich die erlaubten KI-Tools nur im mit der Betreuungskommission vereinbarten Umfang verwendet habe.
In der hier vorliegenden Arbeit habe ich gKI-Systeme wie folgt genutzt:
☐ gar nicht
☐ bei der Ideenfindung
☐ bei der Erstellung der Gliederung
☐ zum Erstellen einzelner Passagen, insgesamt im Umfang von ...% am gesamten Text
☐ zur Entwicklung von Software-Quelltexten/Programm-Code
☐ zur Optimierung oder Umstrukturierung von Software-Quelltexten
✓ zum Korrekturlesen oder Optimieren
☐ bei Übersetzungen
☐ zum Erstellen von Grafiken
✓ Weiteres, nämlich:

Zur Prüfung von Rechtschreibung, Grammatik und Stil wurde ein automatisiertes Sprachkorrekturprogramm (Grammarly) verwendet; inhaltliche Änderungen erfolgten nicht.

Ich versichere, alle Nutzungen vollständig angegeben zu haben. Mir ist bekannt, dass fehlende oder fehlerhafte Angaben als Täuschungsversuch gewertet werden können. Zudem bin ich bereit, auf Nachfrage meine Vorgehensweise (z.B. durch Arbeitsnotizen, Prompts oder mündliche Erläuterungen u.ä.) offen zu legen.
Mir ist bekannt, dass ein Verstoß gegen die genannten Punkte prüfungsrechtliche Konsequenzen hat und insbesondere dazu führen kann, dass die Dissertation abgelehnt wird.

Ort/Datum	Unterschrift Doktorand

Selbstdeklaration bei kumulativen Dissertationen

Konzeption / Planung: Formulierung des grundlegenden wissenschaftlichen Problems, basierend auf bisher unbeantworteten theoretischen Fragestellungen inklusive der Zusammenfassung der generellen Fragen, die anhand von Analysen oder Experimenten/Untersuchungen beantwortbar sind. Planung der Experimente/ Analysen und Formulierung der methodischen Vorgehensweise, inklusive Wahl der Methode und unabhängige methodologische Entwicklung.

Durchführung: Grad der Einbindung in die konkreten Untersuchungen bzw. Analysen.

Manuskripterstellung: Präsentation, Interpretation und Diskussion der erzielten Ergebnisse in Form eines wissenschaftlichen Artikels.

Die Einschätzung des geleisteten Anteils erfolgt mittels Punkteinschätzung von 1 – 100 %.

Für Chapter 4 liegt die Eigenleistung bei 100 % .

Für Chapter 2 liegt die Eigenleistung für

das Konzept / die Planung bei 50%
die Durchführung bei 30 %
die Manuskripterstellung bei 35 %

Für Chapter 3 liegt die Eigenleistung für

das Konzept / die Planung bei 80 %
die Durchführung bei 80 %
die Manuskripterstellung bei 70 %

Die vorliegende Einschätzung in Prozent über die von mir erbrachte Eigenleistung wurde mit den am Artikel beteiligten Koautoren einvernehmlich abgestimmt.

Ort/Datum　　　　　　Unterschrift Doktorand