CESifo WORKING PAPERS

9911
2022

August 2022

Moral Motive Selection in the Lying-Dictator Game

Kai Barron, Robert Stüber, Roel van Veldhuizen

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 9911*

Moral Motive Selection in the Lying-Dictator Game

Abstract

An extensive literature documents that people are willing to sacrifice personal material gain to adhere to a moral motive. Yet, less is known about what happens when moral motives are in conflict. We hypothesize that individuals engage in what we term "motive selection," namely adhering to the moral motive that aligns with their self-interest. We test this hypothesis using a laboratory experiment that induces a conflict between two of the most-studied moral motives: fairness and truth-telling. In line with our hypothesis, our results show that individuals prefer to adhere to the moral motive that is more aligned with their self-interest.

JEL-Codes: C910, D010, D630, D900.

Keywords: motivated reasoning, dictator game, lying game, motives, moral dilemmas.

Kai Barron
WZB
Reichpietschufer 50
Germany – 10785 Berlin
kai.barron@wzb.eu

Robert Stüber
Center for Behavioral Institutional Design
PO Box 129188
UAE - NYU Abu Dhabi
robert.stueber@nyu.edu

Roel van Veldhuizen
Lund University
P.O. Box 7082
Sweden – 22007 Lund
roel.van_veldhuizen@nek.lu.se

August 19, 2022
The authors would like to thank Despoina Alempaki, Vojtech Bartos, Lea Bitters, Lisa Bruttel, Alexander Cappelen, Anastasia Danilov, Martin Dufwenberg, Urs Fischbacher, Tilman Fries, Uri Gneezy, Steffen Huck, Agne Kajackaite, Dorothea Kubler, Max Lobeck, George Loewenstein, Peter Schwardmann, Ivan Soraperra, Bertil Tungodden, and Georg Weizsacker for valuable comments. Financial support from the Deutsche Forschungsgemeinschaft through CRC TRR 190 and from Tamkeen under the NYUAD Research Institute award for Project CG005 is gratefully acknowledged. This study was preregistered at the AEA registry as trial no. AEARCTR-0003617.

1 Introduction

There is ample evidence that many individuals are willing to sacrifice personal gain in order to adhere to a moral principle (or moral motive). These individuals buy fair-trade and organic goods, donate to charity, and refuse to engage in profitable but unethical behavior such as tax evasion or corruption. To model and quantify these moral concerns, previous work has largely relied on stylized settings that make it possible to identify an individual's willingness to pay to comply with a single moral motive. These studies have, for instance, shown that many people are willing to forego material gain in order to reduce inequality (Fehr and Schmidt, 1999) or to avoid lying (e.g., Fischbacher and Föllmi-Heusi, 2013).

Yet, in everyday life, individuals often face scenarios in which multiple moral motives pull them in different directions, with the different moral principles yielding conflicting moral imperatives. For instance, consumers may face a choice between buying a good that is fair-trade or one that is organic. Politicians must often decide between adhering to different fairness principles such as equity or equality. Managers may face a trade-off between improving the working conditions of their workers and safeguarding jobs in the long term. We propose that in such ethical dilemmas, where two or more moral motives are in direct conflict, individuals may choose to adhere to the motive that is most in line with their personal interest. When choosing between different forms of ethical consumption, for example, consumers may favor the one with the lowest price. When deciding between different policy positions, politicians may adhere to the ethical principle that favors their support base and increases their popularity with the current electorate. Finally, managers may focus on goals that increase their own bonus payment.

We term this behavior "motive selection" and study it in a setting that involves a conflict between two of the most well-studied moral motives – truth-telling and fairness. Specifically, we design an experiment that combines key features of two well-known games that are commonly used to study these motives: the dictator game (Forsythe et al., 1994) and Fischbacher and Föllmi-Heusi (2013)'s lying (or cheating) game.[1] As in the lying game, decision makers observe a random number draw and are asked to report its outcome (this random number draw is similar to the dice roll in Fischbacher and Föllmi-Heusi, 2013). Their payment is equal to the number they report. As in the dictator game, any money not claimed by the decision maker is awarded to another participant. We therefore refer to this setting as the LYING-DICTATOR GAME. While truthfully reporting the observed draw may satisfy a desire to be honest, it may also lead to an unfair (unequal)

[1] See, e.g., Shalvi et al. (2011), Gneezy et al. (2018) and Abeler et al. (2019) for a few examples of key recent contributions to this literature.

allocation, generating a conflict between the truth-telling and equality motives.

We implement the game in a laboratory experiment. Our identification strategy relies on the random draw, which creates exogenous variation in the cost of adhering to the truth-telling motive. Consider an individual who wants to behave in a moral way, but cannot simultaneously satisfy both motives. Motive selection predicts that she will choose to satisfy the moral motive that most closely aligns with her personal interest. For low random draws, it is more costly to tell the truth, and motive selection therefore predicts that the individual will adhere to the fairness motive. By contrast, for high random draws, adhering to the fairness motive is more costly, and hence motive selection predicts that individuals will choose to tell the truth instead.

Our results reveal strong evidence that individuals engage in motive selection. Amongst participants with a low random draw (for whom equality is more aligned with self-interest), 47 percent choose the equal division and only 14 percent tell the truth (the rest choose the selfish option of maximizing their payment by keeping the entire endowment). By contrast, participants with a high draw overwhelmingly (75%) choose to tell the truth with only 9 percent choosing to equalize payoffs. In other words, consistent with motive selection, the majority of participants choose to satisfy the moral motive that is most closely aligned with their self-interest. To provide benchmarks for comparing the behavior in the LYING-DICTATOR GAME against, we also run additional sessions in which participants play the baseline dictator and lying game. We find that individuals in the LYING-DICTATOR GAME with *low random draws* behave similarly to individuals playing the classic dictator game, while individuals in the LYING-DICTATOR GAME with *high random draws* behave similarly to individuals in a standard lying game.

Having documented evidence of the existence of motive selection, we proceed to investigating two potential underlying mechanisms. First, participants may have *stable moral preferences* that do not depend on the random draw. According to this hypothesis, the random draw affects the (moral) opportunity cost of adhering to a particular moral motive, but does not affect moral preferences per se (i.e., it does not affect how much utility an individual obtains from complying with a particular moral motive). Second, participants may engage in *motivated reasoning*. According to this perspective, the random draw would directly affect moral preferences by inducing participants to change their utility function to put greater weight on the moral motive that is easier (i.e., cheaper) for them to satisfy.

Our experimental design allows us to distinguish between these two mechanisms in two ways. First, after completing the (first-party) LYING-DICTATOR GAME, participants take part in a spectator version of the same game where their report determines the payments received by two other individuals, but does not affect their own payment. If motive selection is driven by motivated rea-

soning, the random draw in the LYING-DICTATOR GAME should affect the weight attached to a particular moral motive in their utility function. These adjusted utility weights should then spill over to the spectator game leading to the individual again adhering to the moral motive that was aligned with their self-interest in the LYING-DICTATOR GAME. However, we find no evidence of such spillovers in our data. Specifically, neither participants' choices in the spectator lying-dictator game nor their subsequently elicited perceptions of what constitutes socially appropriate behavior indicate any sort of spillovers from their decisions in the (first-party) LYING-DICTATOR GAME. Second, we study motive selection using a model of moral decision making that formalizes equality concerns and lying costs using the models of Fehr and Schmidt (1999) and Abeler et al. (2019), respectively. We calibrate the model using data from our two baseline games and previous work, and compare a version of the model with stable preferences to a version with motivated reasoning. While both versions generate a motive selection effect of similar size, only the stable preferences model can accommodate the reduced frequency of selfish choices we observe in the data compared to the two baseline games. Our results therefore suggest that our data can best be explained by a model with stable preferences.

Our study builds on a large body of work investigating non-selfish motives both theoretically and empirically. Early evidence on *social preferences* showed that people reliably deviate from selfish profit-maximization (e.g., Güth et al., 1982; Kahneman et al., 1986; Forsythe et al., 1994). This inspired models of warm-glow giving and altruism (Andreoni, 1990), reciprocity (Rabin, 1993), inequity aversion (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000) and efficiency concerns (Charness and Rabin, 2002; Engelmann and Strobel, 2004). Recent work examines distributional preferences in settings where income is earned through real-effort tasks (Cappelen et al., 2007; Cappelen et al., 2013; Almås et al., 2020), and shows that social preferences can be changed through early-life interventions (Kosse et al., 2020; Cappelen et al., 2020). We extend this work by studying social preferences in a setting where another moral motive (lying costs) is present.

Evidence regarding *lying aversion* has predominantly been generated using the deception game (Gneezy, 2005; Dreber and Johannesson, 2008; Sutter, 2008; Hurkens and Kartik, 2009; Gneezy et al., 2013) and the lying game (Shalvi et al., 2011; Fischbacher and Föllmi-Heusi, 2013). These studies demonstrate that a large share of people are willing to significantly reduce their earnings in order to avoid telling a lie. Important recent contributions by Dufwenberg and Dufwenberg (2018), Gneezy et al. (2018), and Abeler et al. (2019) suggest that this behavior can best be understood as a psychological cost of lying that depends both on the size of the lie and whether the lie is detected by others. We extend this work by studying lying costs in a setting where social preferences are present as well.

We also contribute to a more limited literature studying the interaction between multiple moral motives. In particular, Cappelen et al. (2007) demonstrate that different individuals may adhere to different fairness motives, and Konow (2000) shows that people may satisfy fairness motives in a self-interested way.[2] We contribute to this work by examining motive selection in a setting where the two moral motives are in different domains (fairness and truth-telling). Further, we do so by designing a framework that provides exogenous variation in the appeal of these motive using a design that can be directly compared to a large number of previous studies.

Finally, our results also contribute to the literature on motivated reasoning and motivated beliefs. These studies show that an individual's self-interest may bias her judgment of what is fair (e.g., Messick and Sentis, 1979; Babcock et al., 1995; Konow, 2000; Gneezy et al., 2019; Amasino et al., 2021), distort her beliefs (e.g., Di Tella et al., 2015; Palma and Xu, 2016; Gneezy et al., 2020), and affect the way that she gathers information (e.g., Babcock et al., 1996; Ambuehl, 2021). We extend this line of work to a setting with two moral motives, and examine whether participants are able to distort the importance of each motive in a motivated way, such that their motivated reasoning justifies self-interested decision making.

The next section describes the experimental design. We present the results in section 3, where we first examine evidence of motive selection and then delve deeper into the mechanisms explaining the evidence. Section 4 concludes.

2 Experimental design and procedures

We study motive selection in a laboratory experiment that we conducted at the WZB-TU laboratory for experimental economics in Berlin in December 2018. We programmed the experiment using zTree (Fischbacher, 2007) and invited participants using ORSEE (Greiner, 2015). Instructions were provided on-screen and are contained in Appendix E. A total of 288 participants took part in the experiments (24 per session).

2.1 Experimental games

Our primary interest in this paper is to investigate behavior in the LYING-DICTATOR GAME, which involves two players: a decision maker (DM) and a recipient. The DM observes a random draw, d, from a uniformly distributed variable with support $\{0, 1, \ldots, 10\}$. She then chooses a number to report, r, which may differ from the randomly drawn number d. The DM receives the value of the

[2] See Capraro and Rand (2018) and Neuber (2021) for recent contributions on this topic.

number reported in Euros, whereas the recipient receives $(10 - r)$. The payoff-maximizing action for the DM is to report $r = 10$; however, she may choose to report a lower number if she has social preferences or a psychological cost of lying.

We compare the decision making in the LYING-DICTATOR GAME to two classic experimental paradigms: the DICTATOR GAME and a version of Fischbacher and Föllmi-Heusi's (2013) LYING GAME. The LYING GAME is similar to the LYING-DICTATOR GAME, except that there is no recipient. That is, any money not taken for oneself is returned to the experimenter. The DICTATOR GAME is also similar to the LYING-DICTATOR GAME, except that there is no random draw. That is, none of the allocations implemented involve telling a lie.

Finally, we also consider behavior in a SPECTATOR LYING-DICTATOR GAME. In this game, the DM also observes a random draw d and subsequently decides what number r to report. The difference is that the number reported now affects the payment of two other players: player A and player B. Player A receives the value of the number reported in Euros, whereas player B receives $(10 - r)$. Hence, the DM's decision does not affect her own payment in this game.

We implemented these four games in two separate treatments, as shown in Table 1. In treatment LYING-DICTATOR, participants play the LYING-DICTATOR GAME followed by the SPECTATOR LYING-DICTATOR GAME. This allows us to observe behavior in the LYING-DICTATOR GAME, and study whether choices in this game spill over to a subsequent game in which the DM's self-interest is removed. In treatment BASELINE, participants play the two canonical games in a random order. This provides us with two benchmark cases against which to compare behavior observed in the LYING-DICTATOR GAME. We will now separately explain these two treatments in greater detail.

Table 1: Overview of experimental treatments

	Session-Level Variation	
	Lying-Dictator	Baseline
Game A	Lying-Dictator Game	Lying Game
Game B	Spectator Lying-Dictator Game	Dictator Game
Order of Games	A then B	Random
N	144	144

Notes: The table describes the experimental variation that we introduce in the study. We had two types of sessions – "Lying-dictator" sessions in which participants played the (first-party) LYING-DICTATOR GAME followed by the SPECTATOR LYING-DICTATOR GAME and "Baseline" sessions in which participants played the standard DICTATOR GAME and LYING GAME, randomly ordered.

2.2 Treatment LYING-DICTATOR

Treatment LYING-DICTATOR consists of three parts. In part 1, participants play the LYING-DICTATOR GAME. In part 2, they play a third-party version of the Lying Dictator Game, namely the SPECTATOR LYING-DICTATOR GAME. In part 3, we elicit the appropriateness of different actions in the SPECTATOR LYING-DICTATOR GAME using the elicitation technique proposed in Krupka and Weber (2013). Finally, participants complete a brief questionnaire and receive their payments.[3]

Upon entering the laboratory, participants were assigned to be one of two player types: Active and Passive players. Active players served as a DM in part 1 and 2. Passive players did not make any decision in part 1, and served as the third party (player A or player B) in part 2. There were always exactly 4 Passive players in each session; the remaining 20 participants in each session were assigned to be Active players. In part 2, two of the Active players' decisions were selected to be relevant for the four Passive players.

2.2.1 Part 1

At the start of the experiment, Active players received the instructions for part 1, the LYING-DICTATOR GAME. Specifically, they were told that they would be presented with a screen containing 11 boxes (see Figure 1).

Figure 1: Stylized depiction of the screen containing the 11 boxes

Notes: Figure 1 provides a stylized depiction of the screen containing the 11 boxes. Participants had to click on one of them, which would then reveal a random number.

Participants were told that they would be asked to click one of the boxes, which would then reveal a random number $d \in \{0, 1, \ldots, 10\}$. They were told they would then move on to another screen and would be asked to report the number they had just seen. They were also told that they would be paid the value of the number reported, r, and any remaining money $10 - r$ would be sent to another participant, the recipient. Other than the presence of the recipient, these procedures are very similar

[3]We preregistered this experimental design at the AEA registry (AEARCTR-0003617), including a pre-analysis plan, power calculation and a detailed discussion of the relationship to the two classical benchmark games; we have reproduced the pre-analysis plan and power analysis in Appendix D.

to Gneezy et al. (2018)'s implementation of the LYING GAME. Note that a key advantage of this procedure is that it allows us to record both the report r and the value of the random draw d.

A key difference between our study and previous work lies in the way participants make their reports. Specifically, we told participants that they could report the number in one of four ways:

1. Tell the truth and report: "The number I saw was [*number seen*]."[4]
2. Equalize payments and report: "The number I saw was 5."
3. Maximize your payment and report: "The number I saw was 10."
4. Maximize the other participant's payment and report: "The number I saw was 0."

This design feature serves two purposes. First, it makes the fact that this is a decision between different motives more salient to participants. We view this as an appealing feature since we are interested in studying situations in which there is a tension between motives, and the salient framing makes this tension explicit. Second, it reduces the number of available reports from 11 (all the possible numbers) to either four or three (depending on whether truth-telling overlaps with one of the other options).[5] This prevents participants from making intermediate choices that do not correspond directly to any of the relevant motives, and sharpens our analysis by making it easier to classify responses as either truth-telling, equalizing or payoff maximization.

We also told participants that after choosing their report they would be asked to provide a brief written explanation for why they chose this report. To maximize the number of participants in the role of DM we used the strategy method (role uncertainty). Specifically, we asked all Active players to make decisions as-if they were the DM, but told them that there was only a 50% chance that their choice was implemented. If not, they would act as recipients for another DM's decision. Roles were only revealed at the end of the experiment. We also told participants that for those whose decisions were implemented, their report (e.g., "The number I saw was 5") would be transmitted to the recipient when payments were revealed at the end of the experiment.

Meanwhile, Passive players were told that they would not make a decision in this part of the experiment. We did present them with the Active players' instructions on their screen. Part 1 of the experiment ended (without providing any feedback) after all Active players had made their reports.

[4]The order of the first two reports, which imply choosing the two moral motives, was randomized between participants, and kept constant across part 1 and part 2.

[5]Even in cases in which truth-telling overlapped with one of the other options, the participants could choose between four options corresponding to the four ways in which the report could be made described above.

2.2.2 Part 2

Active players then received the instructions for part 2, the SPECTATOR LYING-DICTATOR GAME. Similar to part 1, participants were told that they would have to click on a box on their screen to reveal a number, and would then have to report this number in one of four ways:

1. Tell the truth and report: "The number I saw was [*number seen*]."
2. Equalize payments and report: "The number I saw was 5."
3. Maximize player A's payment and report: "The number I saw was 10."
4. Maximize player B's payment and report: "The number I saw was 0."

Relative to part 1, the main difference is that the number reported now does not affect the DM's monetary payoff – it affects the payment of two other players: player A and player B. These are Passive players who do not make decisions in part 2 (although they are able to read the Active players' instructions, as in part 1).

The fact that the DM's part 2 decision is payoff-irrelevant for her implies that it is a pure choice between motives, without her monetary self-interest in play. At the end of the experiment, in each session two of the Active players' decisions are then randomly chosen to determine the payment for two pairs of Passive players. As in part 1, the reports that were implemented were also sent to the recipients (in this case, player A and player B). Part 2 of the experiment ended after all Active players had made their reports.

2.2.3 Part 3

All participants then received the instructions to part 3, the NORM ELICITATION TASK. Following Krupka and Weber (2013), we asked participants to consider the four possible reports made by a hypothetical participant who faced the SPECTATOR LYING-DICTATOR GAME and received a random draw of 8.[6] We then asked participants to rate each of the four possible reports in terms of its "social appropriateness" on a six point scale ranging from "very socially inappropriate" to "very socially appropriate." Participants were told that one of the four reports would be randomly drawn at the end of the experiment, and that they would receive a payment of EUR 2 if their response corresponded to the modal response chosen by participants in the session. Any ties were broken randomly. Part 3 of the experiment ended after all participants had completed all four evaluations.

[6]We chose a random draw of 8, because receiving a draw of 8 implies a clear conflict between truthfully reporting an 8 or choosing to equalize payoffs and reporting a 5.

2.2.4 Questionnaire and payment

Participants then arrived at a payment screen and were informed about which (if any) of their decisions were implemented, as well as the payment they received from each part of the experiment. All participants subsequently went through a brief questionnaire that elicited basic demographics, such as their gender, age, and field of study. In total, each session in this treatment took approximately thirty minutes including payment. The average payment was EUR 12.93 and payments ranged from EUR 7 to EUR 19 (with EUR 7 constituting a fixed payment received by all participants).

2.3 Treatment BASELINE

Treatment BASELINE consists of two parts. In part 1, participants play either the LYING GAME or the DICTATOR GAME, determined at random. In part 2, they play the game they did not play in part 1. At the end of the experiment, one of the two parts is randomly selected for payment. Randomizing the order of the two games allows us to check for order effects. While both games share the same basic structure as the LYING-DICTATOR GAME, the LYING GAME removes the equality motive, while the DICTATOR GAME removes the truth-telling motive.

2.3.1 The LYING GAME

Our main design goal for the LYING GAME was to keep it as similar to the LYING-DICTATOR GAME as possible, while still capturing the key elements characterizing standard lying games observed in the literature. For this purpose, participants again drew a random number by clicking on one of 11 boxes on their screen. They then moved on to another screen where we asked them to report their number in one of four ways:

1. Report: "The number I saw was [*number seen*]."
2. Report: "The number I saw was 5."
3. Report: "The number I saw was 10."
4. Report: "The number I saw was 0."

This implementation ensured that participants could choose between four reports, as in the LYING-DICTATOR GAME. It also kept the decision screens and instructions similar to the LYING-DICTATOR

GAME. The main difference is that the report no longer affects another participant's payment. The purpose of this is to remove the equality motive.[7]

2.3.2 The DICTATOR GAME

Our main design goal for the DICTATOR GAME was also to keep it as similar to the LYING-DICTATOR GAME as possible, while capturing the key characteristics of the standard dictator game. For this purpose, we asked participants to choose between four allocations corresponding to equality, a random draw, payment maximization and payment minimization respectively, just as in the LYING-DICTATOR GAME. As in the LYING-DICTATOR treatment, we asked participants to make decisions as if they were the DM, but told them that there was only a 50% chance that their decision would actually be implemented and otherwise they would act as recipients.

This implementation ensured that participants could choose between 4 allocations, as in the LYING-DICTATOR GAME. It also kept the decision screens and instructions as similar as possible to the LYING-DICTATOR GAME. The key difference is that the random draw was done behind the scenes by the computer, instead of being done explicitly by the participant. The participants did not observe a number that they were asked to report, which ensured that the truth-telling motive could no longer play a role.

2.3.3 Remaining procedures

Upon entering the laboratory, participants were informed that the experiment consisted of two parts and they would be paid for either part 1 or for part 2. Participants then went through the two parts; each part ended only after all participants had finished making their decision in the respective part. Similar to the LYING-DICTATOR GAME, we randomized the order of the first two reports in both parts. We also informed participants that it was possible that two of the possible reports (in the LYING GAME) or allocations (in the DICTATOR GAME) could be identical to ensure they did not think there was a mistake. After part 2 we presented participants with a payment screen notifying participants which of the two games had been selected for payment and displaying their earnings. All participants subsequently went through a brief questionnaire that elicited basic demographics, such as their gender, age, and field of study. In total, each session in this treatment took approximately twenty minutes including payment. The average payment was EUR 13.37 and payments ranged from EUR 7 to EUR 17.

[7] It is worth noting here that this statement assumes, in line with previous work, that the participant's social preferences are not defined with respect to the experimenter as a recipient.

3 Results

In this section we first present evidence that participants engage in moral motive selection in the LYING-DICTATOR GAME. We do this by showing that participants' moral motive choices respond strongly to the random draw they receive. We then explore the extent to which this effect is driven by stable moral preferences or motivated reasoning. Finally, we present additional results on gender, earnings and comparisons to previous work. We provide detailed descriptive statistics in Appendix A. There, we also show that randomization into treatments and of draws within treatments was successful.

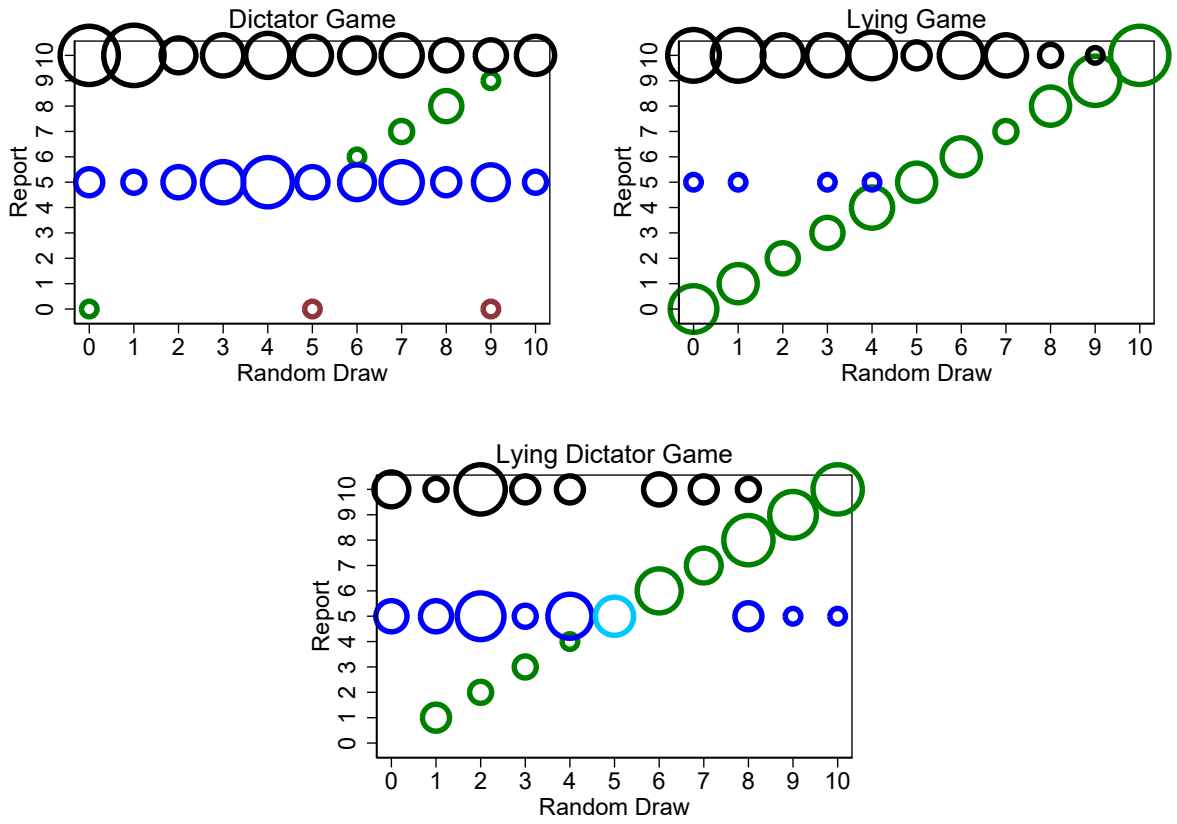
3.1 Motive selection

Figure 2 provides a visual representation of our data. Each panel plots the frequency of observed reports as a function of the random draw; the top two panels present the results of the two baseline games for comparison purposes. The behavior observed in the two baseline games is similar to that documented in previous work. In the DICTATOR GAME, a large fraction of participants choose to either equalize payments (a report of five) or to maximize their own payoff (a report of ten). In the LYING GAME, the majority of participants choose either to tell the truth (the diagonal entries) or to maximize their own payoff (a report of ten). Appendix B.1 provides a more detailed comparison of the behavior observed in these two games to previous studies. Appendix B.2 shows that the order in which the participants took part in the DICTATOR GAME and the LYING GAME do not affect these results.

The lower panel of Figure 2 presents the data of the LYING-DICTATOR GAME. Motive selection predicts that participants tend to pick the motive most closely aligns with their self-interest. Our results are consistent with this prediction. For random draws lower than 5 (LOW draws), almost half (47%) of our participants choose to equalize payoffs, and only 14% choose to tell the truth (diff: 32.76pp, two-tailed test of proportions, $p < 0.001$), see also Figure 3. By contrast, for random draws greater than 5 (HIGH draws), the vast majority (75%) tell the truth and only very few (9%) choose to equalize payments (diff: 66.07pp, two-tailed test of proportions, $p < 0.001$). These results also imply that, consistent with motive selection, moving from a LOW to a HIGH random draw significantly increases the rate of truth-telling (diff.: 61.21pp, two-tailed test of proportions, $p < 0.001$) and significantly decreases the propensity to choose equality (diff.: 37.62pp, two-tailed test of proportions, $p < 0.001$). In other words, participants predominantly select the moral motive that most closely aligns with their self-interest.

Motive selection also predicts that the degree to which behavior in the LYING-DICTATOR

Figure 2: Distributions of choices across the three games

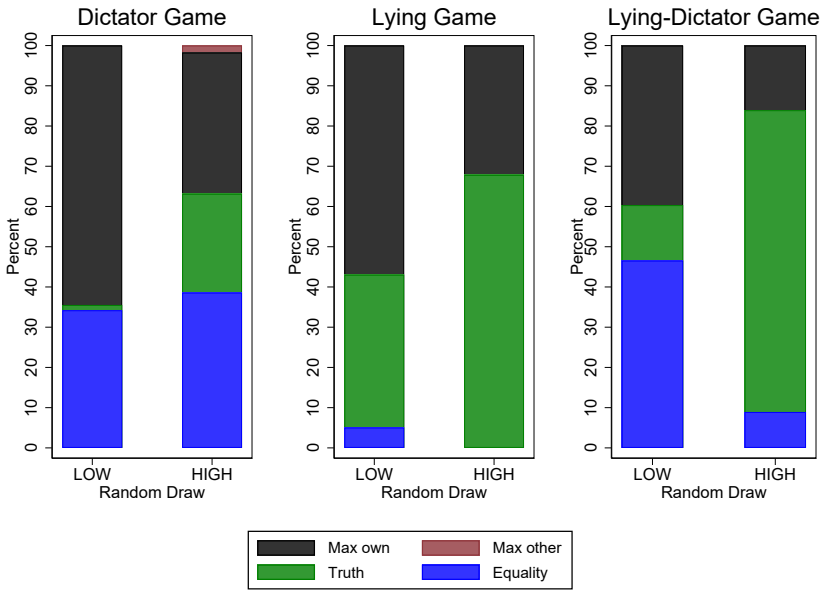


Note: The figure shows the distribution of reports by random draw for each of the three games. Circle sizes correspond to the number of participants.

GAME resembles the two baseline games should depend on the random draw. In particular, behavior after a LOW draw (where the equality motive is least costly) should resemble the DICTATOR-GAME (where equality is the only moral motive). Consistent with this hypothesis, 3 shows that, after a LOW draw, most participants in the LYING-DICTATOR GAME are either selfish or choose equality, similar to the DICTATOR-GAME. Similarly, a HIGH draw leads to mostly selfish or truth-telling choices in the LYING-DICTATOR GAME, similar to the LYING GAME.

This Figure also illustrates that the frequency of selfish behavior is smaller in the LYING-DICTATOR GAME than in the two baseline games. When pooled across LOW and HIGH draws, the fraction of participants choosing the selfish option is 52% and 46% for the DICTATOR GAME and the LYING GAME respectively. By contrast, in the LYING-DICTATOR GAME only 27% of

Figure 3: Motive choices across game types



Notes: The figure shows the distribution of motive choices for a LOW (< 5) draw or a HIGH (> 5) random draw for each of the three games. Note that in the LYING GAME and DICTATOR GAME it was not possible to equalize payments or tell the truth respectively. Instead, for the LYING GAME 'Equality' refers to choosing to take 5 for oneself. For the DICTATOR GAME, 'Truth' refers to choosing to take the randomly drawn number for oneself. Participants who fulfill multiple motives in the lying or dictator game are coded as telling the truth – this can only happen when the random draw is 0 or 10.

participants choose the selfish option ($p < 0.002$ and $p < 0.001$ for the comparison with the DICTATOR GAME and the LYING GAME, respectively).

3.2 Mechanisms of motive selection

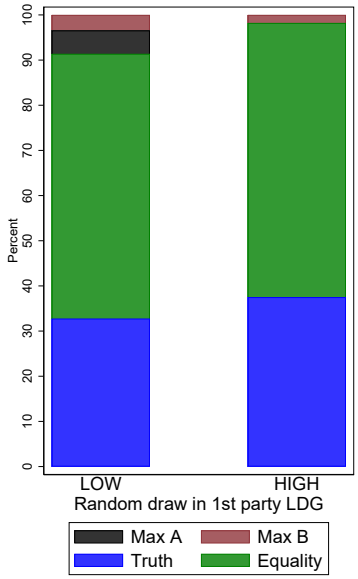
In the previous section, we showed that many participants in our experiment engage in what we called "motive selection," choosing to adhere to the moral motive that is least costly for them to satisfy. In this section, we delve deeper into potential mechanisms generating this effect, focusing on two explanations. First, participants may have *stable moral preferences* that do not depend on the random draw. In this view, the random draw affects the (moral) opportunity cost of adhering to a particular motive, but does not affect moral preferences per se (i.e., it does not affect how much utility an individual obtains from complying with a particular moral motive). Second, participants may also engage in *motivated reasoning*. In this view, the random draw would affect moral preferences by inducing participants to change their utility function to put greater weight on the moral

motive that is easiest (i.e., cheapest) for them to satisfy.

3.2.1 The Spectator Lying-Dictator Game

We are able to distinguish between these two mechanisms experimentally using the data from the SPECTATOR LYING-DICTATOR GAME. In particular, the motivated reasoning explanation posits that participants will put greater weight on the moral motive that is cheapest to satisfy. For example, participants with a high random draw would convince themselves that truth-telling is more important than equality. But if this is true, they should then also be more likely to choose to adhere to the truth-telling motive in the subsequent spectator game, and to attach greater importance to the truth-telling motive in the norm elicitation task. By contrast, if participants have stable moral preferences, the random draw in the LYING DICTATOR GAME should have no spillover effect on the spectator game or the norm elicitation task.

Figure 4: Motive Choices in the SPECTATOR LYING-DICTATOR GAME



Notes: The figure shows the distribution of motive choices in the SPECTATOR LYING-DICTATOR GAME for participants who drew a LOW (< 5) or HIGH (> 5) number in the original LYING-DICTATOR GAME respectively.

Figure 4 presents the choices in the SPECTATOR LYING-DICTATOR GAME as a function of the random draw (LOW or HIGH) in the *first-party* LYING-DICTATOR GAME.[8] The key result

[8] A random draw of 5 in the SPECTATOR LYING-DICTATOR GAME implies that a participant can satisfy both moral motives simultaneously. We keep these observations in our analysis but our results are robust to removing them.

is that the random draw in the LYING-DICTATOR GAME does not have a statistically significant impact on the motive selected in the SPECTATOR LYING DICTATOR GAME. In particular, the truth-telling rate in the spectator game is similar after a LOW draw (33%) and after a HIGH draw (38%; $p = 0.596$, two-sided test of proportions). The equality results are analogous (59% in LOW vs 61% in HIGH, $p = 0.820$).[9] This finding is echoed in the norm-elicitation based on the technique proposed in Krupka and Weber (2013): the random draw in the LYING DICTATOR GAME has no impact on the relative appropriateness ratings for the two motives (two-tailed Mann-Whitney U test, $p = 0.508$).[10] Therefore, we observe no evidence that the motive selection we observe in the first-party LYING-DICTATOR GAME shifts subsequent choices and norm perceptions in the SPECTATOR LYING-DICTATOR GAME in the manner that would be predicted by motivated reasoning.

3.2.2 A model of moral decision making

To further distinguish between and formalize these two explanations, stable moral preferences and motivated reasoning, we present a model of moral decision making that we calibrate based on previous work and the data from our baseline games. We assume that participants have two moral motives (truth-telling and equality), and model these motives using the Abeler et al. (2019) and Fehr and Schmidt (1999) models of truth-telling and social preferences respectively. We incorporate motivated reasoning by allowing participants to (subconsciously) adjust the relative weight given to each motive. The details of the model are contained in Appendix C.

The model has two main implications. First, both versions of the model predict higher truth-telling rates and lower equality rates for high random draws, that is, they predict motive selection. Intuitively, a high random draw reduces the moral opportunity cost of adhering to the truth-telling motive, making it more beneficial to choose it (stable moral preferences) and more beneficial to

[9] Moreover, the distributions do not vary between LOW and HIGH draws (chi-squared test, $p = 0.334$). The results are robust to excluding participants who made selfish choices in part 1 or to only focusing on those participants. In the SPECTATOR LYING-DICTATOR GAME, we observe that 95% of the participants choose either to tell the truth or to divide the endowment equally between the two passive participants. Therefore, with self-interest removed, almost all participants choose to conform with one of the moral motives. Finally, we observe that participants who chose to tell the truth in the LYING-DICTATOR GAME (regardless of their random draw) are somewhat more likely to tell the truth in the spectator version of the game (diff.: 14 pp, $p = 0.098$), but participants who chose equality are only insignificantly more likely to choose equality (diff.: 4 pp, $p = 0.672$).

[10] The random draw in the LYING DICTATOR GAME does also not affect the absolute appropriate ratings of the motive choices. On average, it is more appropriate to tell the truth than to implement equality (Wilcoxon signed-rank test, $p < 0.001$). We find a marginally significant effect of the random draw in a linear regression of the difference in appropriateness rankings between the two motives in which we also control for a dummy for people with a HIGH random draw. This suggests that, controlling for the random draw, there might be a small effect of having a random draw larger than five. The effect is however not very pronounced ($p = 0.094$).

increase the relative weight awarded to the truth-telling motive (motivated reasoning). Second, the stable preference model predicts less selfish behavior in the LYING DICTATOR GAME than in the two baseline games. This is because there are now two moral motives that may induce participants to behave in a non-selfish manner, instead of just one. By contrast, the motivated reasoning model predicts more selfish behavior than both the stable preference model and the baseline games. In particular, when only one of the two motives is strong enough to prevent someone from behaving selfishly, motivated reasoning allows participants to play down that motive in order to behave in a selfish manner. Overall, both models can therefore accommodate the pattern of motive selection that we observe in the data. However, the stable preference model can better explain the reduction in selfish behavior that we observe compared to the baseline games. Taken together, the evidence is therefore more consistent with stable preferences than with the motivated reasoning explanation. This finding is in line with the absence of spillovers observed in the SPECTATOR LYING-DICTATOR GAME described in the previous section.

3.3 Additional insights from the LYING-DICTATOR GAME

In addition to providing evidence of motive selection, the data from the LYING-DICTATOR GAME also reveal several other interesting insights regarding moral decision making. Rather than studying one moral motive in isolation, the LYING-DICTATOR GAME allows us to study how people choose when facing a trade-off between different moral motives while the relative costs associated with adhering to each of the moral motives is varied. We want to emphasize the following four additional observations.

First, it is worth noting that very few participants (9% for LOW draws and 14% for HIGH draws) ever choose to adhere to the more costly moral motive in the LYING-DICTATOR GAME. This suggests that very few participants are deontologists in the sense that they assess which of the two moral motives should be complied with independently of the associated personal financial incentives.[11] This implies that the vast majority of the participants who are complying with a moral motive (i.e., the 47% choosing equality after LOW and the 75% telling the truth after HIGH) would not stick with this moral motive if they received a different draw. In other words, a large share of participants seem willing to treat the moral motives as substitutes, complying with the cheaper one. This also explains why the truth-telling (equality) rate after a LOW (HIGH) draw is lower than observed in the corresponding baseline games.

[11] This finding is in line with Benabou et al. (2020) who document that the share of individuals engaging in deontologically motivated or Kantian rule-based behavior, such as refusing an infinite price to certain moral values, is very small.

Second, the motive choices are substantially more responsive to the random draw in the LYING-DICTATOR GAME than in either of the respective baseline games. To show this, we run regressions in which we regress a dummy for adhering to a moral motive (either truth-telling or equality) as an outcome variable on two indicator variables for (i) a HIGH draw and (ii) the LYING-DICTATOR GAME. We also include the interaction of these two variables.[12] The interaction term of HIGH draw and LYING-DICTATOR GAME is always highly statistically significant implying that the presence of a competing moral motive appears to allow individuals to more easily shift between adhering and not adhering to a particular moral motive as the cost of adherence is varied.[13]

Third, motive selection does not seem to depend on the level of moral behavior in a population. In particular, in the LYING-DICTATOR GAME, women are substantially more likely than men to adhere to a moral motive (12.5% of women choose the selfish option, while 36% of men do, $p = 0.004$).[14] However, both genders engage in motive selection.[15] We view this result as serving as a robustness exercise to our main analysis above, since it shows that the phenomenon of motive selection is observed within two distinct groups that differ in terms of their moral preferences.

Fourth, the decision makers' earnings do not differ significantly across the three games. This is because in the LYING-DICTATOR GAME the reduced propensity to choose the selfish option (which reduces earnings) and the ability to engage in motive selection (which increases earnings) offset one another on average. Importantly, the implication is that in the LYING-DICTATOR GAME the average individual is able to achieve the same earnings, but has a higher propensity of complying with at least one moral principle. Essentially, signaling morality (to oneself or others) becomes cheaper when additional moral motives are introduced, provided one is willing to engage in motive selection.

[12] We therefore run four regressions. Two that compare behavior in the LYING-DICTATOR GAME to behavior in the LYING GAME, with either a binary variable that denotes choosing truth-telling or a a binary variable for equality as the outcome variable. The second pair of regressions replicate this, but compare the LYING-DICTATOR GAME to the DICTATOR GAME.

[13] This comparative static of behavior can be accommodated by both the stable moral preferences and the motivated reasoning models discussed in the previous section.

[14] Whether or not women are more pro-social appears to be context-dependent. In an extensive meta-analysis of the dictator game literature, Engel (2011) finds that women give significantly more. However, in a review of gender differences in preferences across several canonical games, Croson and Gneezy (2009) argue that the evidence is mixed and that women appear more sensitive to experimental cues (see also Exley et al., 2022).

[15] Testing whether the differences in the rate of truth-telling and equality choices are statistically different between LOW and HIGH draws delivers the following results: for women, the rate of truth-telling increases by 59pp and the rate of choosing equality is reduced by 54pp ($p < 0.001$ and $p < 0.001$); for men, the rate of truth-telling increases by 59pp and the rate of choosing equality is reduced by 30pp ($p < 0.001$ and $p = 0.007$).

4 Conclusion

How do people decide what to do in situations where two moral motives are in direct conflict? This paper asks whether individuals use the presence of two conflicting moral motives as an opportunity to pursue their own private interests by satisfying the moral motive that is cheapest to satisfy. We term this behavior "motive selection" and test for its presence using a simple game (the LYING-DICTATOR GAME) that is isomorphic to the classic dictator game, and standard lying game in terms of the mapping from choices into the individual's own payoffs. The only difference between the three games is that we *switch on* or *off* the presence in the choice environment of two moral motives – truthfulness and fairness.

We find that participants in our experiment tend to comply with the more favorable moral motive when more than one is available, consistent with the motive selection hypothesis. We show that this also implies that participants in the LYING-DICTATOR GAME behave as if they are playing a DICTATOR GAME when it is in their private interest to do so, and behave as if they are playing a LYING GAME when this is relatively more advantageous. In addition, having a second moral motive (as in the LYING-DICTATOR GAME) also appears to increase the propensity that an individual chooses to adhere to a moral motive instead of maximizing their own payoff.

Our experimental design allows us to investigate the mechanisms behind motive selection. First, we find no evidence that shifting the moral motive that an individual complies with in the LYING-DICTATOR GAME results in spillovers to their choices in the SPECTATOR LYING-DICTATOR GAME (where the role of self-interest is removed). Nor do we observe any spillovers from the motive selection observed in the LYING-DICTATOR GAME to subsequent perceptions of the appropriacy of complying with different moral motives. These results point towards a limited role of motivated reasoning in this context. Second, we analyze the observed behavior through the lens of two models – one that considers stable moral preferences and and one that allows for motivated reasoning. While both stable moral preferences and motivated reasoning predict motive selection (higher truth-telling rates and lower equality rates for high random draws), the stable preferences model can better explain the reduction in selfish behavior that we observe compared to the two baseline games.

One implication of these results is that the lessons learned from simple decision making contexts, in which there is a tension between monetary gain and satisfying a single moral motive, may not translate directly into more complex decision making contexts where multiple motives are present. In contexts with multiple moral motives, individuals may focus on satisfying at least one of the moral motives, but choose which one they comply with in a selfish way. For instance,

one crucial finding in the literature on lying costs is that the psychological costs of lying seem to be rather large and widespread (Abeler et al., 2014). Our results suggest that individuals may be able to avoid these psychological costs when they are presented with a second moral motive that is "cheaper" to satisfy. Essentially, the introduction of a second moral motive can allow the individual to tell a lie, but still feel that they are behaving morally. Consistent with this, our results show that among participants who are made better off by adhering to another moral motive (equality), we find much lower rates of truth-telling in our LYING-DICTATOR GAME than are commonly found in the literature. We also find lower rates of truth-telling in the LYING-DICTATOR GAME than in our own LYING GAME experiment. Similarly, our participants are much less likely to implement the equal split if the truth-telling motive gives them an excuse not to do so.

One key feature of our experimental design is to provide an abstract setting to study behavior when two conflicting moral motives are present that involves only small deviations from two workhorse games in experimental economics. This allows us, firstly, to increase the complexity of the choice environment in a controlled way – by adding and removing the presence of motives. Secondly, by staying close to the much-studied dictator and lying games, the LYING-DICTATOR GAME game can be analyzed with reference to a wealth of benchmark evidence.

At the same time, the consequences of the type of behavior studied are wide-ranging. Many of the most important decisions in life are complex, and involve an intricate web of competing forces pulling in different directions. From political decision-making to tricky ethical decisions in business to everyday decision-making, complex situations of this nature abound. Our findings suggest that a model of stable moral preferences, which incorporates a fixed cost of violating each moral norm, may be able to accommodate the moral decision making behavior observed in more complex settings. Through this lens, the observed "motive selection" can be thought of as being analogous to a consumption decision, where individuals choose to "buy" compliance with the cheaper moral motive. More research is needed to examine whether these results extend to other settings, but if they do, this indicates that standard models of moral preferences may provide a useful framework for thinking about moral behavior even in more complex settings. Furthermore, our results re-iterate the versatility and usefulness of the standard economics analytical toolbox for providing a lens through which to study a wide array of behaviors.

References

ABELER, J., A. BECKER, AND A. FALK (2014): "Representative evidence on lying costs," *Journal of Public Economics*, 113, 96 – 104.

ABELER, J., D. NOSENZO, AND C. RAYMOND (2019): "Preferences for truth-telling," *Econometrica*, 87, 1115–1153.

ALMÅS, I., A. W. CAPPELEN, AND B. TUNGODDEN (2020): "Cutthroat capitalism versus cuddly socialism: Are Americans more meritocratic and efficiency-seeking than Scandinavians?" *Journal of Political Economy*, 128, 1753–1788.

AMASINO, D., D. D. PACE, AND J. J. VAN DER WEELE (2021): "Fair Shares and Selective Attention," *Tinbergen Institute Discussion Paper 2021-066/I*.

AMBUEHL, S. (2021): "Can Incentives Cause Harm? Tests of Undue Inducement," Unpublished manuscript.

ANDREONI, J. (1990): "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving," *The Economic Journal*, 100, 464.

BABCOCK, L., G. LOEWENSTEIN, S. ISSACHAROFF, AND C. CAMERER (1995): "Biased judgments of fairness in bargaining," *The American Economic Review*, 85, 1337–1343.

BABCOCK, L., X. WANG, AND G. LOEWENSTEIN (1996): "Choosing the Wrong Pond: Social Comparisons in Negotiations That Reflect a Self-Serving Bias*," *The Quarterly Journal of Economics*, 111, 1–19.

BENABOU, R., A. FALK, L. HENKEL, AND J. TIROLE (2020): "Eliciting moral preferences: Theory and experiment," Unpublished manuscript.

BOLTON, G. E. AND A. OCKENFELS (2000): "ERC: A Theory of Equity, Reciprocity, and Competition," *American Economic Review*, 90, 166–193.

CAPPELEN, A. W., A. D. HOLE, E. Ø. SØRENSEN, AND B. TUNGODDEN (2007): "The Pluralism of Fairness Ideals: An Experimental Approach," *American Economic Review*, 97, 818–827.

CAPPELEN, A. W., J. KONOW, E. Ø. SØRENSEN, AND B. TUNGODDEN (2013): "Just Luck: An Experimental Study of Risk-Taking and Fairness," *American Economic Review*, 103, 1398–1413.

CAPPELEN, A. W., J. A. LIST, A. SAMEK, AND B. TUNGODDEN (2020): "The Effect of Early Education on Social Preferences," *Journal of Political Economy*, 128, 2739–2758.

CAPRARO, V. AND D. G. RAND (2018): "Do the right thing: Experimental evidence that preferences for moral behavior, rather than equity or efficiency per se, drive human prosociality," *The Economic Journal*, 99–111.

CHARNESS, G. AND M. RABIN (2002): "Understanding Social Preferences with Simple Tests," *The Quarterly Journal of Economics*, 117, 817–869.

CROSON, R. AND U. GNEEZY (2009): "Gender differences in preferences," *Journal of Economic literature*, 47, 448–74.

DI TELLA, R., R. PEREZ-TRUGLIA, A. BABINO, AND M. SIGMAN (2015): "Conveniently Upset: Avoiding Altruism by Distorting Beliefs about Others' Altruism," *American Economic Review*, 105, 3416–42.

DREBER, A. AND M. JOHANNESSON (2008): "Gender differences in deception," *Economics Letters*, 99, 197–199.

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

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

ENGELMANN, D. AND M. STROBEL (2004): "Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments," *American Economic Review*, 94, 857–869.

EXLEY, C. L., O. P. HAUSER, M. MOORE, J.-H. PEZZUTO, ET AL. (2022): "Beliefs about gender differences in social preferences," University of Exeter Business School Working Paper, No. 22/04.

FEHR, E. AND K. M. SCHMIDT (1999): "A Theory of Fairness, Competition, and Cooperation," *Quarterly Journal of Economics*, 114, 817–868.

FISCHBACHER, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

FISCHBACHER, U. AND F. FÖLLMI-HEUSI (2013): "Lies in disguise—an experimental study on cheating," *Journal of the European Economic Association*, 11, 525–547.

FORSYTHE, R., J. HOROWITZ, N. SAVIN, AND M. SEFTON (1994): "Fairness in Simple Bargaining Experiments," *Games and Economic Behavior*, 6, 347–369.

GNEEZY, U. (2005): "Deception: The role of consequences," *American Economic Review*, 95, 384–394.

GNEEZY, U., A. KAJACKAITE, AND J. SOBEL (2018): "Lying Aversion and the Size of the Lie," *American Economic Review*, 108, 419–53.

GNEEZY, U., B. ROCKENBACH, AND M. SERRA-GARCIA (2013): "Measuring lying aversion," *Journal of Economic Behavior & Organization*, 93, 293 – 300.

GNEEZY, U., S. SACCARDO, M. SERRA-GARCIA, AND R. VAN VELDHUIZEN (2020): "Bribing the Self," *Games and Economic Behavior*, 120, 311–324.

GNEEZY, U., S. SACCARDO, AND R. VAN VELDHUIZEN (2019): "Bribery: Behavioral Drivers of Distorted Decisions," *Journal of the European Economic Association*, 17, 917–946.

GREINER, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, 1, 114–125.

GÜTH, W., R. SCHMITTBERGER, AND B. SCHWARZE (1982): "An experimental analysis of ultimatum bargaining," *Journal of Economic Behavior & Organization*, 3, 367–388.

HURKENS, S. AND N. KARTIK (2009): "Would I lie to you? On social preferences and lying aversion," *Experimental Economics*, 12, 180–192.

KAHNEMAN, D., J. L. KNETSCH, AND R. H. THALER (1986): "Fairness and the Assumptions of Economics," *The Journal of Business*, 59, S285.

KONOW, J. (2000): "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions," *American Economic Review*, 90, 1072–1091.

KOSSE, F., T. DECKERS, P. PINGER, H. SCHILDBERG-HÖRISCH, AND A. FALK (2020): "The Formation of Prosociality: Causal Evidence on the Role of Social Environment," *Journal of Political Economy*, 128, 434–467.

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

MESSICK, D. M. AND K. P. SENTIS (1979): "Fairness and preference," *Journal of Experimental Social Psychology*, 15, 418–434.

NEUBER, T. E. (2021): "Egocentric norm adoption," ECONtribute Discussion Paper No. 116.

PALMA, M. AND Z. XU (2016): *Essays on Fairness Preferences: An Experimental Approach*, chap. Shadow of a Doubt: Moral Excuse in Charitable Giving.

RABIN, M. (1993): "Incorporating Fairness into Game Theory and Economics," *American Economic Review*, 83, 1281–1302.

SHALVI, S., M. J. J. HANDGRAAF, AND C. K. DE DREU (2011): "Ethical Manoeuvring: Why People Avoid Both Major and Minor Lies," *British Journal of Management*, 22, S16–S27.

SUTTER, M. (2008): "Deception Through Telling the Truth?! Experimental Evidence from Individuals and Teams," *The Economic Journal*, 119, 47–60.

APPENDICES

A Descriptive statistics

A total of 288 participants took part in the experiments spread evenly across the two treatments. Out of the 144 participants in the Lying-Dictator treatment, 120 were Active players and we will subsequently focusing on these. We include all 144 participants of the Baseline treatment.

Overall, 63% of participants were men. Ninety-eight percent were students and the average age was 22.7. Participants were approximately in their fifth semester. Participants studied various fields, most study engineering and related fields (46%), business/economics (17%), and maths/information science (13%). As shown in Panel A of Table A.1 these characteristics were balanced across the two treatments: There is no statistically significant difference in gender composition, age, and length of study (semester) between treatments. The participants are somewhat more experienced with lying experiments in treatment Baseline, the difference is however not significant at the five-percent level. Also, participants do not vary in their fields of studies between treatments (not reported).

Panel B of A.1 shows that in the three games, participants were equally likely to observe a high or low random draw. This is crucial, because we do not fix the random draws after observing them for either of the treatments and because whether a participant observes a high or low random draw provides the exogenous variation exploited to investigate whether the participants engage in motive selection.

Table A.1: Descriptive Statistics

Panel A. Randomization of covariates across treatments

	Overall	Lying-Dictator	Baseline	Δ
Male	0.633	0.600	0.660	0.316
	(0.483)	(0.492)	(0.475)	
Age	22.66	22.27	22.99	0.162
	(4.157)	(3.380)	(4.695)	
Semester	4.712	4.542	4.854	0.522
	(3.934)	(3.678)	(4.143)	
Experience with	1.640	1.383	1.854	0.051
lying games	(1.957)	(2.216)	(1.689)	
Observations	264	120	144	

Panel B. Randomization of draws within treatments

	Overall	Lying-Dictator Game	Lying Game	Dictator Game	χ^2
HIGH	0.414	0.467	0.389	0.396	0.379
	(0.493)	(0.501)	(0.489)	(0.491)	
LOW	0.522	0.483	0.549	0.528	0.564
	(0.500)	(0.502)	(0.499)	(0.501)	
Five	0.064	0.050	0.062	0.076	0.681
	(0.245)	(0.219)	(0.243)	(0.267)	
Observations	408	120	144	144	

Note: Panel A. contains means and standard errors of several participant characteristics pooled for all participants (column 1) and for both treatments (column 2 and 3), as well as the *p*-values for comparisons between columns 2 and 3 based on a two-sided test of proportions for the categorical variable (male) and a two-sided t-tests for the other variables (column 4). Panel B. displays the proportion of HIGH random draws, LOW random draws, and random draws equal to five for all games (column 1) and for the three games separately (column 2 to 4), as well as the *p*-value from Pearson's chi-squared tests. For the Lying-Dictator Game only Active players are included.

B Robustness checks for Lying Game and Dictator Game

B.1 Comparison to literature

In this section, we compare our results in the two baseline games to previous work. This allows us to test whether specific features of our design and population lead to behavior that differs from previous work or is similar. Overall, our results replicate the main stylized facts established in previous work.

B.1.1 Lying Game

We first compare the results from our LYING GAME to the findings of Gneezy et al. (2018), whose implementation of the lying game is similar to ours in that they also ask participants to click on a box on the screen and to report the observed number.[16] Crucially, this implementation of the lying game allows the researcher to observe lying at the individual level. Our results are presented in Figure B.1.

The comparison yields the following findings: First, in line with Gneezy et al. (2018), we find that reported numbers are significantly higher than observed numbers (two-tailed t-test, $p < 0.001$), because a substantial fraction of participants lie. Second, as in Gneezy et al. (2018) we find a significant negative correlation between the number observed and the probability of lying (Spearman's rho = -0.329, $p < 0.001$). Hence, participants who observe a lower number are more likely to lie. Third, in line with Gneezy et al. (2018), conditional on lying, we find no correlation between the observed and reported number (Spearman's rho = 0.114, p=0.349). Fourth, as in Gneezy et al. (2018) we find that most participants who lie, lie maximally and report a 10. In our data, of the participants who lie, 94% of participants report a 10 (while 6% report a 5).[17] This statistic is robust to restricting attention to the participants who observe a LOW draw and lie – of these 49 participants, only four (8%) lie partially and report a 5, while 45 (92%) lie maximally and

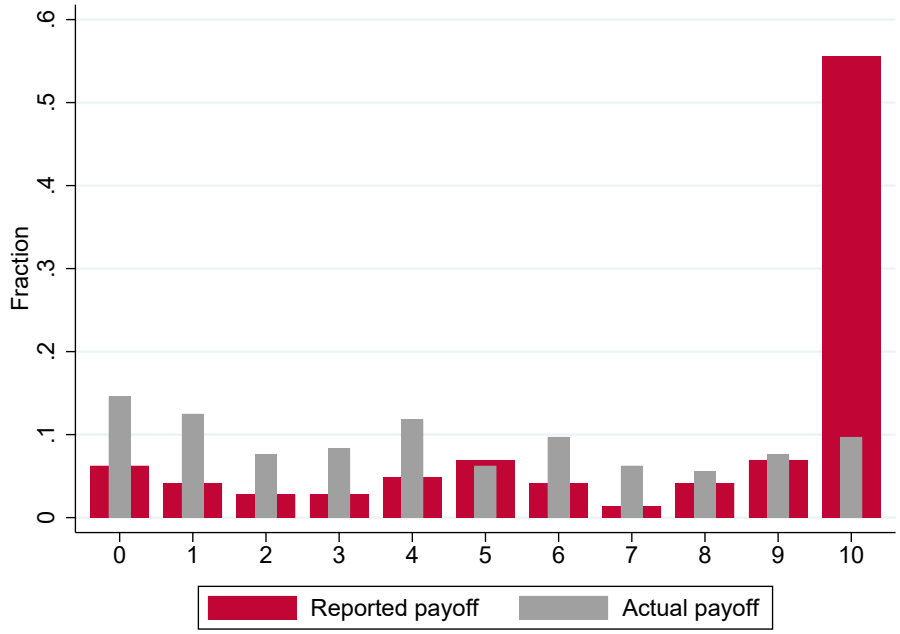
[16] One difference is that in Gneezy et al. (2018) participants can observe a number between one and 10, while in our study the random number varies between zero and 10 (see the *Numbers treatment* of their *Observed Game*). In our setting it is important to allow participants to be able to choose zero, as it facilitates symmetry of the own-other allocation choices, and enhances comparability with our DICTATOR GAME. A second difference is that we focus directly on the tension between motives by restricting each participant's choice set to four items in each of the three games.

[17] In the Gneezy et al. (2018) Numbers treatment, 68% of participants who lie, lie maximally. This number increases to 80% in their Numbers Mixed treatment, and to 91% in their Words treatment. We view these results as being in line with ours. In particular, since participants in our experiment can only lie to 0, 5 or to 10, and the most of those who do not lie maximally in Gneezy et al. (2018), lie to 9 or 8. In our paper and in all three treatments in Gneezy et al. (2018), fewer than 10% of those who lie, report a 5 or less. In both papers, there is almost no evidence of downward lying.

report a 10. Hence, there is almost no partial lying, perhaps because the random draw is observed.

Thus, the findings in our LYING GAME resemble the pattern found in Gneezy et al. (2018). One difference between our results and the ones reported in Gneezy et al. (2018) is that we observe more lying. The fraction of participants who lie is 49% which stands in contrast to the 26% of participants that lie in the Numbers treatment of Gneezy et al. (2018). One potential reason for this is our restriction of the choice set to four items. This element of the experimental implementation may legitimize lying slightly and induce a higher rate of lying. Note, however, that since this is held constant across our games, we do not view it as a major concern for treatment comparisons.

Figure B.1: Distributions of reports in the Lying Game



Note: The figure shows the numbers reported by participants in the Lying Game (red bars) and the actual number they observed (gray bars).

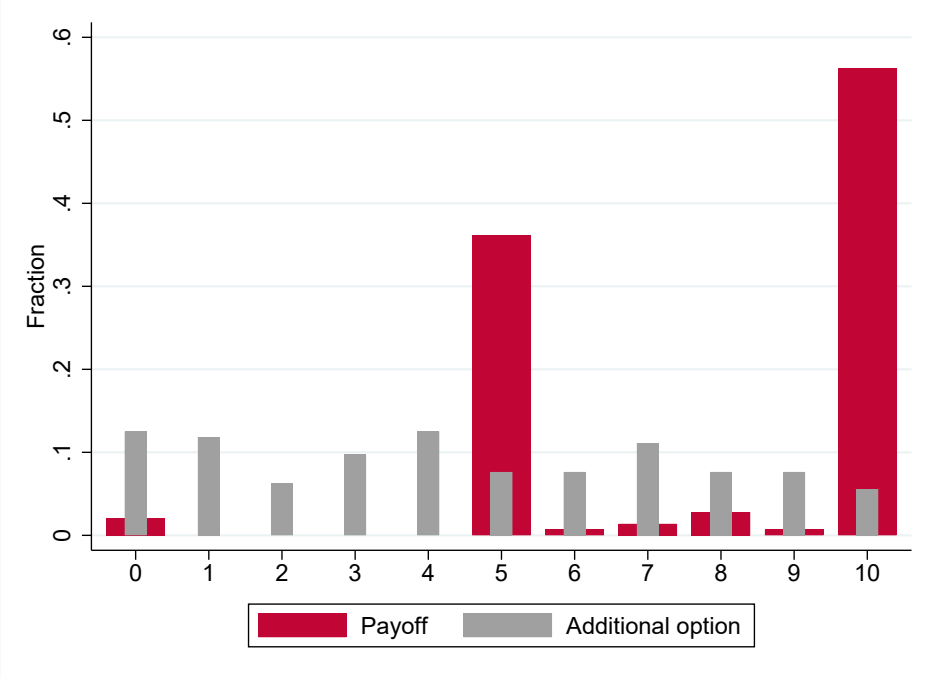
B.1.2 Dictator Game

We also compare our findings in the DICTATOR GAME to previous work. Given the large pre-existing literature studying the game, we view the most appropriate benchmark for comparison as being the aggregated behavior observed across most published studies. We therefore compare our findings to the results reported in the meta-analysis of Engel (2011). More specifically, we compare our findings to the distribution of individual giving rates based on 328 treatments and 20,813 observations reported in this paper (see his Figure 2).

Our distribution of dictator choices is shown in Figure B.2 and shares the following features with the one reported in Engel (2011). Firstly, the modal choice for participants is to retain 100% of the endowment for themselves. Secondly, the second most commonly chosen option is to give 50% of the endowment to the recipient and to keep 50% for oneself. Thirdly, higher numbers are chosen more often than lower numbers, that is, more participants choose to retain 60% to 90% of the endowment than 10% to 40% of the endowment. Fourth, the fraction of participants that give everything to the recipient is slightly larger than the fraction giving 60% to 90%. Hence, the distribution of DICTATOR GAME-giving obtained in our experiment replicates the major stylized facts established in various earlier studies and summarized in Engel (2011).

Specific to our setup is the finding that the fraction of participants who choose to retain 100% is larger than the fraction reported in Engel (56% vs. 36% in Engel, 2011) and the fraction of participants who choose to retain more than 50% but less than 100% smaller (in total 34% of participants in Engel (2011), but only 5% in our study). This concentration of choices on retaining 50%, and 100% is likely a mechanical consequence of our four-option design, which implies that only one out of every 11 participants in our experiment can choose each of the outcomes other than 0, 5, and 10€. Hence, a participant that, for instance, would like to keep 7, 8, or 9€ for themselves might end up choosing 10€. A similar logic can explain the increase in the proportion of participants choosing to retain 50% (36% vs. 17% in Engel, 2011). Yet, the mean contribution in our setup is comparable to the mean from all reported or constructed means in Engel (2011) (21% vs. 28% in Engel, 2011). Hence we conclude that our DICTATOR GAME-results replicate those in the literature.

Figure B.2: Distributions of choices in the Dictator Game



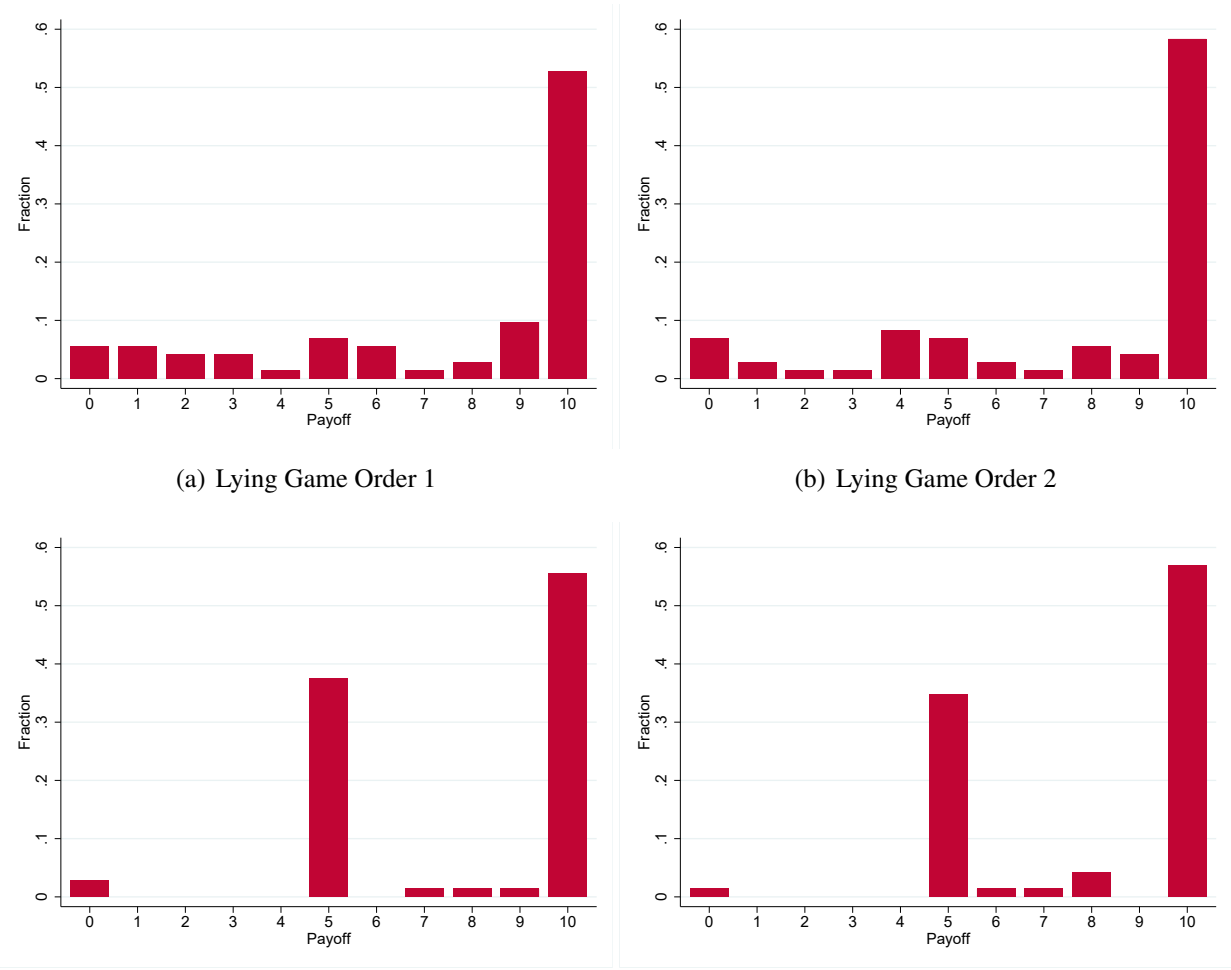
Note: The figure shows the distribution chosen by participants in the Dictator Game (red bars) and the number they could have chosen in addition to 0, 5, and 10 € (gray bars).

B.2 Order effects

In treatment BASELINE we vary the order of the LYING GAME and the DICTATOR GAME. We check whether our results with respect to the two games are robust to order effects. We label the order "Order 1" if participants first participated in the LYING GAME and "Order 2" if they first made a choice in the DICTATOR GAME.

As illustrated in Figure B.3, we do not find any evidence of order effects, that is, the order in which participants participate in the two games does not influence our findings. In particular, for the LYING GAME the reported distributions are not significantly different from one another between the two orders (Kolmogorov-Smirnov test, $p = 0.995$, Mann-Whitney U test, $p = 0.609$). For the DICTATOR GAME, we also find that the distribution of choices made are not statistically different from one another (Kolmogorov-Smirnov test, $p = 1$; Mann-Whitney U test, $p = 0.722$). Equally, average choices in the DICTATOR GAME are 7.94 € for participants who first play the DICTATOR GAME and 7.76 € for participants who play the LYING GAME first (two-tailed t-test, $p = 0.682$), and in the LYING GAME they are 7.47 € for participants who first play the LYING GAME and 7.69 € for participants who play the DICTATOR GAME first (t-test, $p = 0.696$).

Figure B.3: Distributions for the first and second choice made in Treatment BASELINE



(a) Lying Game Order 1

(b) Lying Game Order 2

(c) Dictator Game Order 1

(d) Dictator Game Order 2

Note: The figure shows the (unconditional) distributions for the Lying Game and the Dictator Game both for participants who participated first in the Lying Game (order 1) and who played the Dictator Game first (order 2).

C Modeling mechanisms of motive selection

The main result of our experiment is that many participants engage in what we call "motive selection," choosing to adhere to the moral motive that is less costly for them to satisfy. In this section, we present a theoretical framework to better understand and distinguish between two potential explanations generating this effect. The first explanation maintains that participants have *stable moral preferences* that do not depend on the random draw. In this view, the random draw affects the opportunity cost of adhering to a particular motive, but does not affect moral preferences per se. The second explanation instead proposes that participants engage in *motivated reasoning*. In this view, the random draw affects moral preferences directly by inducing participants to change their utility function to put greater weight on the moral motive that is easiest (i.e., cheapest) for them to satisfy.

C.1 Theoretical framework

To formalize this distinction, let us consider an agent i who is choosing which number $x_i \in [0, 10]$ to report in the LYING-DICTATOR GAME. While the agent benefits financially from reporting a higher number, she may face a moral cost for doing so based on violating one of two moral motives: equality and truth-telling. We parameterize these moral costs using Fehr and Schmidt (1999)'s model of inequity aversion for the equality component and Abeler et al. (2019)'s model of preferences for truth-telling for the truth-telling component respectively. Agent i's utility function is:

$$\begin{aligned} U(x_i) &= x_i - [\alpha_1 I(x_i \neq \bar{x}_i)] - [\alpha_2(x_i - x_j)I(x_i > x_j) - \alpha_3(x_j - x_i)I(x_i < x_j)] \\ &= x_i - [\text{Lying Costs}] - [\text{Inequity Aversion}] \end{aligned} \tag{1}$$

Here, $\bar{x}_i$ is the random number drawn and x_j is the payment for the other player. The three parameters α_1, α_2, and α_3 reflect the individual's cost of lying, aheadness aversion, and behindness aversion, respectively. The social preference component is identical to Fehr and Schmidt (1999)'s model of inequity aversion. The lying cost component is based on Abeler et al. (2019); the fact that all lies are detected by the experimenter simplifies the lying cost component of the model, and allows us to interpret α_1 as capturing both a fixed costs of lying and a disutility from the proba-

bility of being seen as a liar.[18] Note that we assume that the two moral components are additively separable for simplicity.

While the previous model assumes stable moral preferences, we can extend it to incorporate motivated reasoning as follows:

$$\begin{aligned} U(x_i) &= x_i - \gamma[\alpha_1 I(x_i \neq \bar{x}_i)] - (2-\gamma)[\alpha_2(x_i - x_j)I(x_i > x_j) - \alpha_3(x_j - x_i)I(x_i < x_j)] - c(\gamma) \\ &= x_i - \gamma * \text{Lying Costs} - (2-\gamma) * \text{Inequity Aversion} - c(\gamma) \end{aligned} \tag{2}$$

Intuitively, the agent can pay a psychological cost $c(\gamma)$ to adjust the relative weight $\gamma \in [0, 2]$ put on each of the motives. In other words, the model assumes that the agent can use motivated reasoning to emphasize the motive that agrees with an action, and downplay the motive that disagree with it. Note that the model holds the total weight awarded to both motives constant; only the relative weight can be changed through motivated reasoning. For simplicity, we assume that $c(\gamma) \geq 0$, $c(1) = 0$ and that $c(\gamma)$ is increasing in the absolute distance from $\gamma = 1$. This also implies that the stable preference model is a special case of the motivated reasoning model where $\gamma = 1$.

In the next sections, we will calibrate both versions of the model and compare their predictions to the data from the LYING-DICTATOR GAME. In doing so, it is useful to note that the model can also be applied to the two single-moral motive games. Assuming that lying aversion plays no role in the DICTATOR GAME and assuming no motivated reasoning ($\gamma = 1$), equations 1 and 2 for the DICTATOR GAME translate to:

$$U(x_i) = x_i - \alpha_2(x_i - x_j)I(x_i > x_j) - \alpha_3(x_j - x_i)I(x_i < x_j) \tag{3}$$

For the LYING GAME, the corresponding expression is:

$$U(x_i) = x_i - \alpha_1 I(x_i \neq \bar{x}_i) \tag{4}$$

[18] Abeler et al. (2019)'s preferred model models lying cost as a combination of a fixed psychological cost of telling a lie and an image cost that depends on the probability with which the experimenter expects a particular report to be a lie. In our experiment, however, all lies are detected by the experimenter, allowing us to capture both costs with a single parameter.

C.2 Calibrating the model

In order to differentiate between stable preferences and motivated reasoning, we calibrate our model based on previous work and the data from our baseline games. For the LYING GAME, the calibration in Abeler et al. (2019) implies that $\alpha_1 \sim U[3, 15]$.[19] Since this leads to less lying than we observe in the baseline game, we instead assume that one third of the population has $\alpha_1 = 0$, and the other two thirds have $\alpha_1 \sim U[3, 15]$. Under this assumption, 56% of participants are expected to lie, which accommodates the share observed in the data well. For the DICTATOR GAME, our model predicts that participants with $\alpha_2 > 0.5$ will choose to equalize payments, but puts no restriction on the behindness aversion parameter α_3. Instead, we therefore take the distribution of social preference-types proposed by Fehr and Schmidt (1999). In particular, we assume that 30% of participants are selfish ($\alpha_2 = 0, \alpha_3 = 0$), 30% have $\alpha_2 = 0.25$ and $\alpha_3 = 0.5$, 30% have $\alpha_2 = 0.6$ and $\alpha_3 = 1$, and 10% have $\alpha_2 = 0.6$ and $\alpha_3 = 4$. This parameter combination predicts that 40% of participants will choose to equalize payments in the DICTATOR GAME, and can therefore accommodate our baseline data fairly well (see Figure B.2).

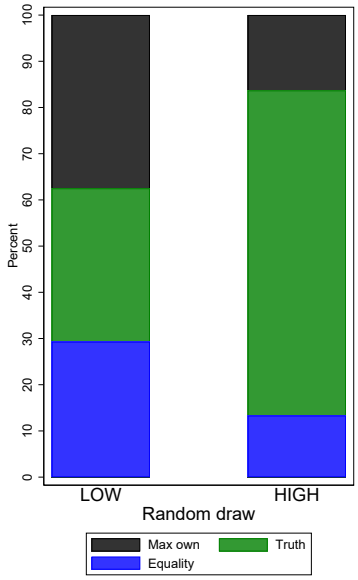
C.2.1 Stable Moral Preferences

We can use these calibrated parameter values to generate predictions for the LYING-DICTATOR GAME, starting with the case of stable moral preferences ($\gamma = 1$). For this purpose, we assume that social preferences and lying costs are independent; the results for correlated preference terms are similar and presented in the discussion section below. The results of the calibration (Figure C.1) show that the stable preference model already generates a HIGH/LOW effect that goes in the same direction as our results. That is, the calibrated model predicts that even under stable moral preferences, participants will engage in motive selection. Intuitively, the appeal for the truth-telling option is greater for high draws because the monetary payoff is larger and because in the Fehr-Schmidt model participants are more averse to disadvantageous inequality than advantageous inequality. By contrast, equality and payoff maximization are equally appealing for both types of draws. This generates the HIGH/LOW effect that we see in the figure. The calibrated model also predicts a lower frequency of selfish choices than in the two baseline games, something we also observe in our data (see Figure 3).

However, the HIGH/LOW effect is somewhat smaller than the effect we find in our data.

[19] Abeler et al. (2019) model lying costs using $U(x_i) = x_i - cI(x_i \neq \bar{x}_i) - \theta\Lambda(x_i)$, where c is a fixed cost of lying, $\Lambda(x_i)$ is the probability that the experimenter thinks a report of x_i as a lie and θ is an image cost. Abeler et al. (2019)'s calibrated model assumes that $c = 3$ and $\theta \sim U[0, 12]$. Since all lies are detected in our experiment, $\Lambda(x_i) = 1$ whenever $x_i \neq \bar{x}_i$. Taken together, this implies that $\alpha_1 = c + \theta \sim U[3, 15]$.

Figure C.1: Stable Preference Model Predictions



Note: This figure shows the predicted fraction of participants maximizing their payment, equalizing payments and telling the truth for HIGH (> 5) and LOW (< 5) draws respectively, for the stable preference model. The figure assumes that lying costs and social preferences are independent.

Whereas in our data, having a HIGH draw reduces the tendency to choose equality by 38pp and increases the tendency to tell the truth by 61pp; the model only predicts effects of 16pp and 38pp, respectively. This suggests that allowing for motivated reasoning may improve the fit of the model by generating a larger motive selection effect.

C.2.2 Motivated Reasoning

We allow for motivated reasoning by assuming that participants are able to adjust the relative weight for the motives (γ) to maximize their payment x_i. One way to formalize this intuition is to use a dual-self model in which a 'doer' chooses which number to report and a 'planner' chooses the value γ to induce the doer to choose the financially most beneficial action (i.e., to maximize x_i). Their corresponding utilities may look as follows.

$$\begin{aligned} U_d(x_i) &= x_i - \gamma[\alpha_1 I(x_i \neq \bar{x}_i)] - (2-\gamma)[\alpha_2(x_i - x_j)I(x_i > x_j) - \alpha_3(x_j - x_i)I(x_i < x_j)] \\ U_p(x_i) &= x_i - c(\gamma) \end{aligned} \tag{5}$$

For simplicity, we assume that adjusting the weights is costless (so that $c(y) = 0$). As a first step, it is easy to see that the doer would never choose to maximize the other participant's payment. This implies that that the doer is choosing between three options: maximizing her own payoff, equalizing payments and telling the truth.[20] Let us first consider the trade-off between maximizing ($x_i = 10$) and equalizing ($x_i = 5$) payments. Assuming that neither of the two options coincides with telling the truth (i.e., that $\bar{x}_i \notin \{5, 10\}$), we obtain:

$$U_d(10) - U_d(5) = 5 - 10(2 - \gamma)\alpha_2 \tag{6}$$

This equation tells us that participants with limited social preferences ($\alpha_2 \leq 0.25$) will prefer to maximize their own payment even if all weight is put on the social preferences component ($\gamma = 0$). Since setting $\gamma = 0$ removes all lying cost, this implies that whenever the doer has weak social preferences ($\alpha_2 \leq 0.25$), the planner will be able to ensure that she receives the maximum payment by setting $\gamma = 0$. Given the distribution of social preferences in Fehr and Schmidt (1999), this would imply that 60% of participants would be able to maximize their payment regardless of the random draw.

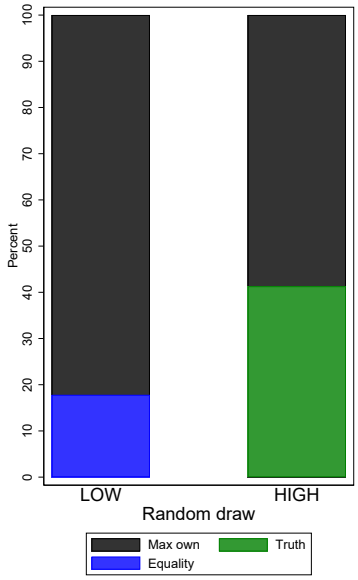
The remaining 40% of the population have $\alpha_2 = 0.6$. Among this group, for participants with zero lying cost, the planner can set $\gamma = 2$ and maximize her payment that way. Among participants with a positive lying cost, we can differentiate between low and high draws. In case of a high draw, the worst possible outcome is equality, so the planner should try to ensure that the doer's utility of obtaining 10 exceeds the utility of equality. The previous equation tells us that this will be the case whenever $\gamma \geq \frac{7}{6}$. However, for these participants, the planner will also want to minimize the appeal of the truth-telling option by setting γ as low as possible. Taken together, this implies that these participants will set $\gamma = \frac{7}{6}$. For the distribution of lying costs we assume, this implies that among participants with a high draw and $\alpha_2 = 0.6$, one third (those with zero lying cost) will set $\gamma = 2$ and maximize their payment. The remainder will set $\gamma = \frac{7}{6}$ and choose to tell the truth.[21]

In case of a low draw, participants with zero lying cost can still set $\gamma = 2$ and maximize their payment. The remaining participants can always guarantee a payment of 5 by setting $\gamma = 0$. However, they may be able to achieve the maximum payment of 10 by setting $\gamma = \frac{7}{6}$, which will

[20] This follows from the fact that participants in the experiment could only choose between these four options. However, it can be shown that even in a setting where participants could report any number $x_i \in [0, 10]$, no participant would ever report any other number than $\bar{x}_i$, 5, or 10.

[21] Formally, $\gamma = \frac{7}{6}$ implies that $U_d(5) = U_d(10) = 5 - \frac{7}{6}\alpha_1 \leq 5 = U_d(\bar{x}_i)$.

Figure C.2: Motivated Reasoning Model Predictions



Note: This figure shows the predicted fraction of participants maximizing their payment, equalizing payments and telling the truth for HIGH (> 5) and LOW (< 5) draws respectively, for the motivated reasoning model. The figure assumes that lying costs and social preferences are independent.

be optimal as long as this value of γ does not induce the doer to tell the truth, that is, as long as:

$$U_d(10) - U_d(\bar{x}_i) = (5 - \gamma\alpha_1) - (\bar{x}_i - (2 - \gamma)\alpha_3(10 - 2\bar{x}_i)) > 0 \tag{7}$$

This equation illustrates that the relative utility of maximizing payment is decreasing in the lying cost α_1, increasing in the behindness aversion parameter α_3 (either 1 or 4 for participants with $\alpha_2 = 0.6$), and (since $(2 - \gamma)\alpha_3 > 0.5$) decreasing in the random draw. If $U_d(10) \geq U_d(\bar{x}_i)$ when $\gamma = \frac{7}{6}$, the planner will set $\gamma = \frac{7}{6}$ and the doer will choose to maximize their payment. In other cases, the planner will set $\gamma = 0$ and obtain a payment of 5.

Figure C.2 present the model predictions, maintaining the assumption that social preferences and lying costs are independent. Like the model with stable preferences, the motivated reasoning model also predicts a HIGH/LOW effect. In fact, the predicted effect (18pp for equality and 41pp for truth-telling) is nearly identical to the stable preference model prediction (16pp and 39pp respectively). That is, allowing for motivated reasoning does not increase the predicted size of the motive selection effect. Where the two models do generate different predictions is in the frequency of selfish choices, which is much higher under motivated reasoning (59% and 82% under LOW

and HIGH draws, respectively) than under stable preferences (16% and 38%). Intuitively, this is because motivated reasoning allows the planner to set $\gamma \neq 1$ in a self-serving way, which helps make payoff maximization the utility maximizing case in a greater number of cases. The other difference lies in the frequency of the more costly motive, which is chosen with positive probability under stable preferences but never chosen under motivated reasoning.

C.3 Discussion

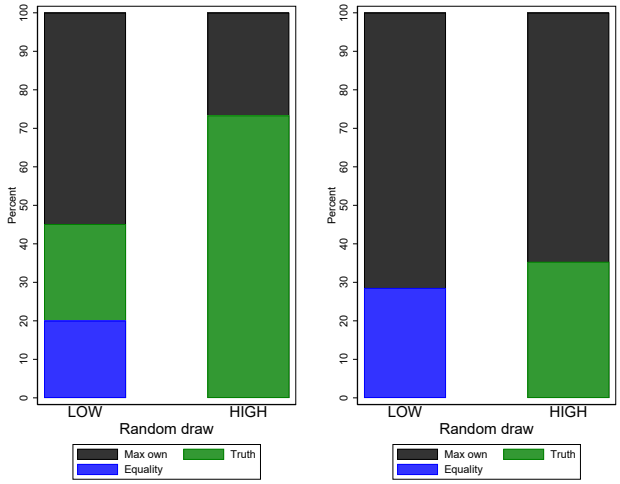
What do the results of these calibrations tell us about the ability of the two models to accommodate the patterns we observe in our data? First, we saw that both models generate a near-identical motive selection (HIGH/LOW) effect. Therefore, when it comes to the HIGH/LOW effect, both models are equally consistent with the data. In contrast, when it comes to the fraction of selfish behavior, the stable preference model tracks the results of the experiment more closely than the model with motivated reasoning. When it comes to the frequency choosing the more costly motive, the results of the experiment fall somewhere in between the two models.

Why does motivated reasoning not increase the motive selection (HIGH/LOW) effect in our model? Intuitively, motivated reasoning increases the scope for motive selection by allowing participants to adjust γ to increase the appeal of the less costly motive. However, motivated reasoning also decreases the scope for the HIGH/LOW effect by increasing the scope for selfish behavior. In particular, motivated reasoning allows participants to shift γ to decrease the overall appeal of moral behavior. For example, a participant with a random draw of 0 might choose to emphasize the truth-telling motive as a way to justify choosing the selfish option over equality. This allows a large portion of participants to behave selfishly regardless of the random draw, reducing the scope for finding a HIGH/LOW effect. These two effects cancel out in our calibration, leading to a near-zero effect of motivated reasoning on the motive selection (HIGH/LOW) effect overall.

While all of these results are generated using a specific parametric functional form, the main comparative statics appear to be robust to alternative specifications. For example, one potential reason that motivated reasoning may have increased the scope for selfish behavior is that our calibration assumes that social preferences and lying costs are independent. Yet, when we instead assume that social preferences and lying costs are perfectly rank-correlated, we obtain very similar results as shown in Figure C.3. In particular, the HIGH/LOW effect is similar for both models, and the frequency of selfish behavior is still higher under motivated reasoning.

The models considered so far generate a motive selection (HIGH/LOW) effect, but the predicted effect tends to be smaller than the one we observe in the data. This may be the result of the specific

Figure C.3: Predictions under Correlated Preferences



(a) Stable Preferences (b) Motivated Reasoning

Note: This figure shows the predicted fraction of participants maximizing their payment, equalizing payments and telling the truth for HIGH and LOW draws respectively. The figure assumes that lying costs and social preferences are perfectly rank correlated. The left and right panel present the results for the stable preference and motivated reasoning models respectively.

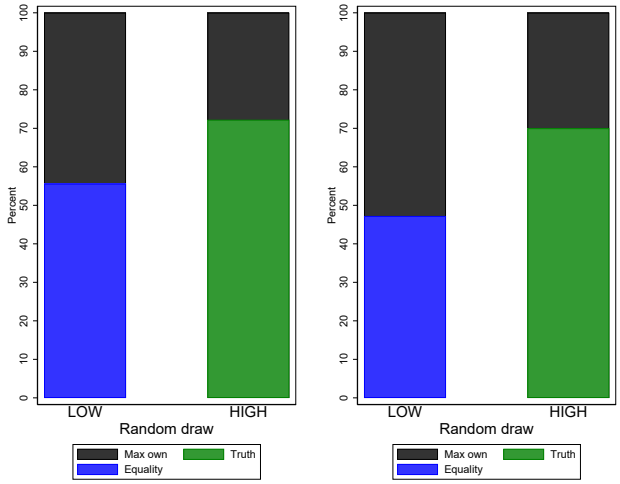
way we calibrated our model. In particular, we can also consider the following simple model:

$$U(x_i) = x_i - \gamma[\alpha I(x_i \neq \bar{x}_i)] - (2-\gamma)[\alpha I(x_i \neq 5)] \tag{8}$$

This specification assumes that participants pay a fixed moral cost of α for any motive they fail to adhere to. Similar to α_1 in the previous model, the assumption that one third of the population has $\alpha = 0$, and the other two third has $\alpha \sim U[3, 15]$ accommodates the data from the baseline games well.

Figure C.4 shows that this model fits the data very well under the assumption of stable preferences ($\gamma = 1$). In particular, it predicts a HIGH/LOW effect (56pp for equality and 72pp for truth-telling) that is more in line with our data (38pp and 61pp) and larger than the previous models (16pp and 38pp for the main specification). It also predicts the rate of selfish choices quite well (44% and 28% after LOW and HIGH choices respectively; this is 40% and 16% in the data). Note that because lying costs are assumed to be identical to inequality costs, participants will always choose the motive that is cheapest for them to satisfy. Motivated reasoning slightly increases the scope for selfish behavior by allowing participants to lower the weight on the more attractive

Figure C.4: Alternative Model Predictions



(a) Stable Preferences (b) Motivated Reasoning

Note: This figure shows the predicted fraction of participants maximizing their payment, equalizing payments and telling the truth for HIGH and LOW draws respectively. The figure assumes that lying costs and social preferences are perfectly rank correlated. The left and right panel present the results for the stable preference and motivated reasoning models respectively.

motive as a way to increase the relative appeal of the selfish option. In keeping with previous models, however, motivated reasoning has only a minimal effect on the size of the HIGH/LOW effect, slightly reducing it. Overall, the simple alternative model fits the data very well, and further confirms that allowing for motivated reasoning is unlikely to affect the size of the HIGH/LOW effect (motive selection) in a major way.[22]

Finally, it is worth noting that the predictions of the stable preferences and motivated reasoning model also differ in terms of the motive selected *conditional on not being selfish*. In particular, under motive selection, non-selfish participants will always select the cheaper motive (truth-telling for high draws, equality for low draws). Under stable preferences, however, participants may still select the more expensive motive. In particular, the main model with independent (perfectly cor-

[22]To behave selfishly in this model, it needs to be true that both $U(10) \geq U(5)$ and $U(10) \geq U(\bar{x}_i)$, which implies that both $\gamma \geq \frac{2\alpha-5}{\alpha}$ and $\gamma \leq \frac{10-\bar{x}_i}{\alpha}$. Under stable preferences $\gamma = 1$, and agents will be able to behave selfishly if $\alpha \leq \min(5, (10 - \bar{x}_i))$. Under motivated reasoning, the planner can adjust the value of γ to increase the range of moral costs α for which both equations are satisfied. In particular, the two inequalities jointly imply that agents will be able to behave selfishly as long as $\alpha \leq 7.5 - 0.5\bar{x}_i$. Compared to the stable preference case, the planner will inflate the weight put on the more costly motive in order to induce the doer to act selfishly when the stable preference. However, in practice this will only be beneficial for a small range of α, which explains why the two models yield similar results in this case.

related) moral costs predicts that 47% (56%) of non-selfish participants choose to tell the truth for a LOW draw, and 16% (0%) may choose equality for a HIGH draw. In the data, the corresponding fractions are 23% and 11% respectively, which lies in between the two model predictions. The specific predictions are also sensitive to assumptions on the self-deception cost $c(\gamma)$ and the specific functional form. When self-deception costs are positive, non-selfish participants may no longer always choose the cheaper motive even under motivated reasoning. When using a different functional form, such as equation (8), even the stable preference model may predict that all non-selfish participants will choose the cheaper motive.

All in all, the analysis in this section reveals three results. First, the stable preference model already generates a HIGH/LOW effect in the LYING-DICTATOR-GAME. Second, allowing for motivated reasoning does not consistently increase the size of this effect (though it does increase the effect among non-selfish participants). Third, allowing for motivated reasoning increases the frequency of selfish behavior. This implies that we cannot reliably differentiate between the two models through looking at the HIGH/LOW effect, but can do so by comparing the rate of selfish behavior to the baseline games. Contrary to the predictions of the motivated reasoning model, but in line with the stable preference model, we find that the frequency of selfish behavior is reduced in the LYING-DICTATOR GAME compared to the two baseline tames. Taken together, our experimental data are therefore more consistent with a model with stable preferences than a model with motivated reasoning.

D Pre-Registration

In this section, we reprint the pre-analysis plan and power calculation that we included in our pre-registration.[23] For the pre-analysis plan, we also note where the results of the pre-registered specifications can be found in the paper. Note that we only pre-registered the LYING-DICTATOR treatment. The primary purpose of the baseline treatment is to serve as a comparison for the patterns observed in the LYING-DICTATOR GAME.

D.1 Pre-Analysis Plan

This pre-analysis plan consists of two parts. First, we describe the statistical tests we intend to conduct in the paper. Second, we provide some further theoretical background for our main hypothesis, i.e., that participants with a LOW draw in part 1 (S1<5) are more likely to choose equality in part 2 than those with a high draw (S1>5), and less likely to choose truth-telling.

D.1.1 Preliminaries

1. We define the LOW group and HIGH group as those participants who received a low draw (S1<5) and a high draw (S1>5) in Part 1 respectively. Unless otherwise indicated, all analysis will be carried out only on the participants in these two groups (i.e., we will not use the data from Passive players or those with a random draw of S1=5).

2. As a manipulation test, we run two-sided tests of proportions to check whether HIGH participants were indeed less likely to report equality and more likely to choose truth-telling than LOW participants. This allows us to see whether our manipulation successfully induced different behavior in the two groups.

3. As a further manipulation test, we will also show regressions of a dummy for choosing a motive on a dummy for the HIGH group and the signal from part 1. We will run a separate regression for both the truth-telling and equality motive. This allows us to see whether the exact draw S1 affected the propensity to select a given motive conditional on being in the HIGH or LOW group.

[23] The study was preregistered at the AEA registry as trial no. AEARCTR-0003617 and can be found here: https://www.socialscienceregistry.org/trials/3617. The preregistration also included theoretical considerations that partially overlap but not coincide with the model proposed in the paper, and that we therefore not reproduce here.

4. We compute the fraction of participants in each group who chose neither equality nor truth in Part 2. If this fraction is non-negligible (>10%), our main analysis will separately report the results for truth-telling and equality. If not, the results for truth-telling and equality are likely to be identical. Hence we will report only the results for equality in the main text, and only note those cases in which the results for equality and truth-telling differ using e.g., a footnote in the text.

Authors' notes: We divided the sample as specified in (1); (2) is the main test for motive selection presented in Section 3.1. The regression analysis in (3) yielded a significant effect for the HIGH group and no significant effect for the exact random draw S1 and we therefore omitted it for brevity. We refer to the robustness checks mentioned in (4) when discussing the results of the SPECTATOR LYING-DICTATOR GAME in Section 3.2.

D.1.2 Main Analysis

5. Two-sided test of proportions testing whether the fraction of participants choosing equality in the HIGH group in Part 2 differs significantly from the LOW group. This allows us to test our main hypothesis, i.e., that participant with a HIGH draw in Part 1 are less likely to report equality in Part 2.

6. Two-sided test of proportions testing whether the fraction of participants choosing truth-telling in the HIGH group in Part 2 differs significantly from the LOW group. This allows us to test our main hypothesis, i.e., that participant with a HIGH draw in Part 1 are more likely to report the truth in Part 2.

Authors' notes: These tests are presented in Section 3.2 as a way to test whether motive selection is driven by motivated reasoning.

D.1.3 Robustness Analysis

7. Linear regression of a dummy for choosing a given motive in Part 2 on a dummy for the HIGH group plus the signal received in Part 1. This allows us to test whether the exact signal received matters even after controlling for whether it was a LOW or HIGH signal. We will run separate regressions for the two main motives (equality and truth-telling).

8. Re-do the main analysis (points 5 and 6) separately for participants who chose the selfish motive in part 1, and those who did not. This allows us to see whether the main effect we

observe in points 5 and 6 comes from those who did not choose the selfish motive in part 1, those who did, or both.

Author's notes: These tests are reported in footnotes presented in Section 3.2.

D.1.4 Additional Analysis

9. Two-sided Mann-Whitney test investigating whether the difference between the appropriateness scores for equality and truth-telling in part 3 differs between the HIGH and LOW group. In addition, we run a linear regression of the difference between the appropriateness score for equality and the appropriateness score for truth-telling on a dummy for the HIGH group plus the signal received in Part 1. This analysis allows us to test whether members of the two groups differ in the relative appropriateness rankings for the truth-telling and equality motive.

Authors' notes: We include both these tests in Section 3.2.

D.2 Power calculations

Our key hypothesis is that the motive selected in Part 2 (M2) depends on the random number drawn in Part 1 (S1). Specifically, we hypothesize that participants with a low draw (S1<5; LOW group) are more likely to choose the equality motive than participants with a high draw (S1>5; HIGH), and vice versa for the truth-telling motive.

To determine the sample size required to test this hypothesis, we require a quantitative prediction of the effect size we may expect in the experiment. We do this by combining the classic model of inequality aversion model of Fehr and Schmidt (1999) and the more recent model of lying costs of Abeler et al. (2019; equation 1):

$$U(r,s,\delta) = r + \delta * [\kappa_1\theta_1|r-s| - \kappa_2\theta_2 I_{LIE} * \frac{r_{max} - r_{min}}{2}] + (1-\delta) * [\alpha * (10-2r)I_{DIS} - \beta * (2r-10)I_{ADV}]$$

The first term in this utility function (r) represents the participant's monetary payoff, which is equal to their report. The second term is the lying cost function of Abeler et al. (2019). Here, θ_1 and θ_2 are parameters that measure the direct lying cost and social image cost of lying respectively. I_{LIE} is an indicator for any report that deviates from the signal s. Since the experimenter observes both r and s, we assume that all lies are detected and are equally harmful to the social image.

Following Abeler et al. (2019), we further assume that θ_1 and θ_2 are independently uniformly distributed on [0,1]. Also following Abeler et al. (2019), the calibration parameters κ_1 and κ_2 are set to 3 and 4 respectively, and from the design of our experiment r_{max} and r_{min} are equal to 10 and 0 respectively. The third term in the utility function is the inequality aversion model of Fehr and Schmidt (1999). We assume that the parameters for the utility loss from advantageous β and disadvantageous α utility are distributed exactly as in Fehr and Schmidt (1999, Table III). Finally, $\delta \in$ [0,1] determines the weight given to lying cost and social preferences respectively.

In generating predictions for part 1 of the experiment, our key initial assumption is that δ=0 for (S1<5) and δ=1 for (S1>5). Intuitively, this assumption implies that agents with a low draw will choose to adhere to the moral motive that allows them to pick a higher number than their initial draw (equality). By contrast, agents with a high draw will adhere to the moral motive that will allow them to report their high draw without feeling bad for doing so (the truth-telling motive).

Taken together, these assumptions lead to the following predictions for Part 1. Among low draws, it is easy to show that 40% of the population will choose equality and report 5, with the remainder choosing payoff maximization (i.e., report 10). This follows directly from the parameters calibrated by Fehr and Schmidt (1999), for whom 40% of the population have $\alpha \geq 1$ and would therefore be willing to reduce their own earnings in order to achieve equality. No intermediate reports are observed as the result of the linearity of the model. Among high draws, 98% of the population chooses to tell the truth, the remaining 2% chooses to maximize their payoff. This follows directly from the parameters calibrated by Abeler et al. (2019). Intuitively, lying costs are thought to be very high, and the benefit to lying given a high draw is small, meaning that only agents with very low draws of both θ_1 and θ_2 decide to deviate from truth-telling. Partial lies are not observed because there is no benefit to lying partially given that image costs of lying do not depend on the specific lie that is told.

To generate predictions for part 2 of the experiment, we then assume that agents who chose either truth-telling or equality in part 1 will choose the same motive in part 2. We also assume that those who chose payoff maximization in part 1 will randomize between equality and truth-telling in part 2, and are equally likely to choose either motive. If this is true, 99% of agents with high draws in part 1 will choose truth-telling in part 2, compared to 30% of agents with low draws; for equality the corresponding percentages are 1% and 70%.

Given these predictions, we are able to determine the minimum sample size required in each of the two groups (HIGH draw or LOW draw) to obtain a power of .80 to detect a difference in the proportion of participants in the HIGH and LOW group choosing the truth-telling motive, respectively, using a test of proportions. The results of these calculations are presented in Table D.1

for the truth-telling motive; the results for the equality motive are identical given the assumption that all participants choose either equality or truth-telling in part 2. In addition to presenting the baseline results, Table D.1 also presents the minimum sample size required to detect somewhat smaller effects (e.g., if not all participants choose δ in the self-serving way assumed in our derivations above) as well as the effect of random noise, that is, participants randomizing between the equality and truth-telling report in part 2 (e.g., because they care little about the outcome given the lack of a personal stake). For example, the fourth entry in the second row shows that 35 people are needed in both groups to obtain a power of .80 if the rate of truth-telling in part 2 is equal to 40% and 90% in each group respectively, and 30% of people in the experiment fully randomize between motives in part 2.

Table D.1: Sample Size Calculations

	Sample Size (Power 0.80)					
	Noise: 0%	Noise: 10%	Noise: 20%	Noise: 30%	Noise: 40%	Noise: 50%
Truth-telling						
99% vs 30%	9	12	15	19	26	37
90% vs 40%	17	21	27	35	48	68
80% vs 50%	45	56	72	94	127	183

Notes: Each row presents the minimum sample size required per group (HIGH or LOW draw) to obtain a power of 0.80 given the effect size specified in the first column. The columns vary the fraction of the population that is assumed to randomize between the equality and truth-telling motive.

Overall, Table D.1 illustrates that greater noise and smaller effect sizes require a larger sample size. Note that while we used a specific parametric specification to generate these predictions, we obtain similar results if we replace the theoretical predictions with the average behavior observed in previous experiments, if we use different models of social preferences, if we assume that participants can freely choose their δ to maximize their own utility, or if we assume that δ is assumed to be weakly increasing in s (although the exact point predictions will differ slightly).

To determine the appropriate sample size for our study, we conservatively assume that the effect size may be smaller than the predictions of the model, and there may be some noise as well. Specifically, we aim to collect 50 observations in both the LOW group and the HIGH group. Keeping in mind that random draws of 5 are in neither group, we would therefore need 110 participants to achieve our target sample size of 50 participants in each of the two groups. In addition, we require at least 20% more participants to serve as Passive players and a few more participants to guard against unbalanced sampling due to random variation, bringing us up to a total target sample size of 140 participants.

E Experimental instructions

This appendix contains the experimental instructions for treatment Lying-Dictator (Appendix E.1) and treatment Baseline, that is, the LYING GAME and DICTATOR GAME) Appendix E.2). Parts in *italic* were not shown to the participants. Original instructions were in German.

E.1 Instructions for the LYING-DICTATOR GAME

Here we present the instructions for the LYING-DICTATOR GAME, separately for Active players and Passive players.

E.1.1 Active players

Screen 1: **General instructions**

Welcome to today's experiment

For showing up today, you will receive a 5 € show-up fee. In addition to this amount, you can earn money during the experiment. The amount of money you earn will depend on the decisions that you and/or other participants make.

During this experiment you are not allowed to use electronic devices, or to communicate with other participants. Please only use the programs provided for the experiment. If you have a question, please remain seated and raise your hand. We will then come to you and will answer your question in private. Please do not ask your question out loud. If you violate these rules you will be excluded from the experiment and the payment.

As with all experiments in this experimental laboratory, there is no deception: That means that whenever we tell you that something will happen if you make a certain choice, then that thing will happen if you make that choice.

This experiment consists of three parts. You will receive the instructions for the first part on the next screen. The instructions to part two and three will be provided after you have finished part one and two, respectively. For simplicity, we only use the male terminology; of course any explanation is also meant to include women.

Part 1: Lying-Dictator Game
Screen 2:
No participation.

Screen 3: **Instructions Part One**

In part one of the experiment you will see 11 boxes with numbers hidden in them. The numbers in the boxes are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 and they are placed in a random order. We will ask you to click on one box. Once you click on the box, you will see a number that we ask you to remember and later report to us. The number you report to us determines how much money you and one other participant in the experiment will be paid.

The payments corresponding to each number are as follows:

Number Reported	You receive	Other participant receives
0	0€	10€
1	1€	9€
2	2€	8€
3	3€	7€
4	4€	6€
5	5€	5€
6	6€	4€
7	7€	3€
8	8€	2€
9	9€	1€
10	10€	0€

After clicking on a box and seeing a number, you can make your report in one of four ways:[24]

- Tell the truth and report: "The number I saw was [number seen]."
- Equalize payments and report: "The number I saw was 5."
- Maximize your payment and report: "The number I saw was 10."
- Maximize the other participant's payment and report: "The number I saw was 0."

Your report ("The number I saw was ...") will be sent to the other participant. In addition, after you have made your decision, you will be asked to briefly write an explanation for why you chose your selected report.

The other participant will also simultaneously complete the same task. After you and the other participant have both completed the task, either your decision or the other participant's decision will be implemented. There is a 50% chance that your decision will be implemented and a 50% chance

[24]The order of the first two options was randomized and then kept constant throughout the instructions.

that the other participant's decision will be implemented. If your decision will be implemented your payment is determined by your own decision only. If the decision of the other participant will be implemented, your payment is determined only by the other participant's decision.[25] At the end of the experiment, you will be informed whose decision is implemented. If your decision is implemented, your report ("The number I saw was ...") will be sent to the other participant and will determine your payment and the payment of the other participant. If their decision is implemented, their report will be sent to you and will determine both payments.

If you have any questions please raise your hand. We will come to you and answer them in private.

Screen 4:

Please click on one of the boxes:

Screen 5:

Please choose the report you would like to make. On the right you see the payment implied by each report.

Your report	You receive	Other participant receives
○ Tell the truth and report: The number I saw was d.	d €	(10-d) €
○ Equalize payments and report: The number I saw was 5.	5 €	5 €
○ Maximize your payment and report: The number I saw was 10.	10 €	0 €
○ Maximize the other participant's payment and report: The number I saw was 0.	0 €	10 €

Screen 6:

You chose the option: Tell the truth and report: The number I saw was d.

Please briefly explain why you have chosen this option.

Screen 7: **End Part One**

The first part of the experiment has finished. The second part will begin shortly. Please press "Continue."

Part 2: Spectator Lying-Dictator Game

Screen 8:

No participation.

[25] The last two sentences were not included in the instructions of the pilot session.

Screen 9: **Instructions Part Two**

In this part of the experiment you will again see 11 boxes with numbers hidden in them. As before, the numbers in the boxes are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 and they are again placed in a random order. Again, we will ask you to click on one box. Once you click on the box, you will see a number that we ask you to remember and later report to us.

The difference is that this time, the number you report to us **does not determine** your payment. Instead, it determines the payment of **two other participants** who are also in the lab right now, Participant A and Participant B. Note that these two participants **did not** make a decision in the first part of the experiment and **will not** make a decision in the second part of the experiment (i.e. this part), but are only reading through the instructions. In other words, these participants' earnings depend on your decision.

The payments corresponding to each number are as follows:

Number Reported	Participant A receives	Participant B receives
0	0€	10€
1	1€	9€
2	2€	8€
3	3€	7€
4	4€	6€
5	5€	5€
6	6€	4€
7	7€	3€
8	8€	2€
9	9€	1€
10	10€	0€

After clicking on a box and seeing a number, you can make your report in one of four ways:

- Tell the truth and report: "The number I saw was [number seen]."
- Equalize payments and report: "The number I saw was 5."
- Maximize Participant A's payment and report: "The number I saw was 10."
- Maximize Participant B's payment and report: "The number I saw was 0."

A total of four participants in the lab have been randomly selected not to make a decision in the first and second part of the experiment. Two of these participants are in the role of participants A

and two in the role of participants B. While you are completing part 1 and part 2 of the experiment, these participants are asked to read the instructions, they however are not making any decisions themselves.

After all those participants who do make decisions in part 1 and 2 (such as you), have made their decisions, two of them will be randomly selected and their decisions will be implemented. This means that if you are selected, your decision is implemented and your report will be sent to two participants in the lab – a Participant A and a Participant B – and will determine their payments.

If you have any questions please raise your hand. We will come to you and answer them in private.

Screen 10:

Please click on one of the boxes:

Screen 11:

Please choose the report you would like to make. On the right you see the payment implied by each report.

Your report	Participant A receives	Participant B receives
○ Tell the truth and report: The number I saw was d.	d €	(10-d) €
○ Equalize payments and report: The number I saw was 5.	5 €	5 €
○ Maximize Participant A's payment and report: The number I saw was 10.	10 €	0 €
○ Maximize Participant B's payment and report: The number I saw was 0.	0 €	10 €

Screen 12: **End Part Two**

The second part of the experiment has finished. The third part will begin shortly. Please press "Continue."

Part 3: Norm elicitation.

Screen 13: **Instructions Part Three**

In this part of the experiment you will be asked to rate the social appropriateness of each of the four choice options that a hypothetical participant could have chosen between in the previous part of the experiment (part two). More precisely, please imagine a participant in the experiment who clicks on a box and observes an **8**. This participant can therefore choose between the following four options:

- Tell the truth and report: "The number I saw was 8."
- Equalize payments and report: "The number I saw was 5."
- Maximize Participant A's payment and report: "The number I saw was 10."
- Maximize Participant B's payment and report: "The number I saw was 0."

Imagine the payments corresponding to each number are again as follows:

Number Reported	Participant A receives	Participant B receives
0	0€	10€
1	1€	9€
2	2€	8€
3	3€	7€
4	4€	6€
5	5€	5€
6	6€	4€
7	7€	3€
8	8€	2€
9	9€	1€
10	10€	0€

Please indicate for each possible option (1 to 4), whether choosing this option is "very socially inappropriate," "socially inappropriate," "somewhat socially inappropriate," "somewhat socially appropriate," "socially appropriate," or "very socially appropriate." By **socially appropriate**, we mean behavior that most people agree is the **"correct"** or **"ethical"** thing to do.

At the end of the experiment, the computer will randomly select one of the four possible options. For the option selected, we will determine which rating **was chosen by the most other participants** in the laboratory. If you chose the rating that was also chosen by most other participants, you will receive a payment of **2€**. If your rating does not correspond to the rating chosen by most other participants, you will receive **0€**.

For example, if the third option ("Maximize Participant B's payment and report: The number I saw was 0") is randomly selected and you choose "very socially inappropriate," then you will receive 2€ if this is also the rating that was chosen by most other participants in the lab. If other participants instead chose one of the other ratings ("socially inappropriate," "somewhat socially inappropriate," "somewhat socially appropriate," "socially appropriate," or "very socially appropriate") more often then you receive no additional payment. If several ratings are chosen equally often, one of the most frequently chosen ratings will be randomly selected.

If you have any questions please raise your hand. We will come to you and answer them in private.

Screen 14:

For each possible option, please indicate whether you believe choosing that option is "very socially inappropriate," "socially inappropriate," "somewhat socially inappropriate," "somewhat socially appropriate," "socially appropriate," or "very socially appropriate." One of the four options will be randomly selected. For the option selected, we will determine which rating most other participants in the laboratory chose. If you chose the rating that was also chosen by most other participants, you receive 2 €.

Recall that the participant clicks on the box and observes an **8**.

1. Tell the truth and report: "The number I saw was 8."

○ Very socially inappropriate	○ Socially inappropriate	○ Somewhat socially inappropriate	○ Somewhat socially appropriate	○ Socially appropriate	○ Very socially appropriate

2. Equalize payments and report: "The number I saw was 5."

○ Very socially inappropriate	○ Socially inappropriate	○ Somewhat socially inappropriate	○ Somewhat socially appropriate	○ Socially appropriate	○ Very socially appropriate

3. Maximize Participant A's payment and report: "The number I saw was 10."

○ Very socially inappropriate	○ Socially inappropriate	○ Somewhat socially inappropriate	○ Somewhat socially appropriate	○ Socially appropriate	○ Very socially appropriate

4. Maximize Participant B's payment and report: "The number I saw was 0."

○ Very socially inappropriate	○ Socially inappropriate	○ Somewhat socially inappropriate	○ Somewhat socially appropriate	○ Socially appropriate	○ Very socially appropriate

Screen 15: **Your income today**

Thank you! For completing the experiment you receive an additional payment of 2 €. Your income today hence consists of the following:

Show-up fee: 5 €

If randomly selected as decision maker: In part 1 of the experiment, the computer randomly selected your decision to be implemented. You reported: "The number I saw was x." Hence you earn x.

If not randomly selected as decision maker: In part 1 of the experiment, the computer randomly

selected the other participant's decision to be implemented. The other participant reported: "The number I saw was y." Hence you earn: $(10 - y)$.

In part 2 of the experiment you reported: "The number I saw was z." *If not randomly selected:* Your decision was not selected to be implemented. *If randomly selected:* Your decision was randomly selected to be implemented. Hence Participant A will receive z and Participant B will receive $(10 - z)$.

In part 3 of the experiment option q was randomly chosen for payment. You indicated that you believed this report to be "p." This was (not) the rating chosen by most other participants. Hence you earn s.

Payment for completing the experiment: 2 €.

Your total payment is therefore: t.

Please answer some questions on the next screens[26], before you receive your payment.

Questionnaire

Screen 16:[27] **Questionnaire**

Do you remember which option ("Tell the truth...," "Equalize payments...," "Maximize your payment...," "Maximize the other participant's payment...,") you have chosen in part 1? If yes please enter the option. If not, please enter "I don't remember."

Do you remember which option ("Tell the truth...," "Equalize payments...," "Maximize Participant A's payment...," "Maximize Participant B's payment...,") you have chosen in part 2? If yes please enter the option. If not, please enter "I don't remember."

Please now consider the decisions that you made in part 1 and 2 simultaneously. Maybe you have chosen the same option in both part 1 and 2. However, maybe you have chosen a different option in part 1 and 2. Please briefly explain your choice (i.e., if you have chosen the same option in part 1 and 2, why did you choose the same option in both parts and if you have chosen a different option in part 1 and 2, why did you choose a different option?)

Screen 17: **Questionnaire**

Please answer the following questions:

Please indicate your gender.

How old are you (in years)?

[26] In the pilot this read: "on the next screen."

[27] Not part of the instructions in the pilot.

What do you study? In case you do not study, please enter your current occupation.

In which term are you? In case you do not study, please enter a 0.

How often have you previously participated in an experiment in this laboratory? In case you have never participated, please enter a 0.

How often have you previously participated in a similar experiment? By similar experiment we mean for instance an experiment in which you had to roll a die and report the outcome of the die roll or something similar.

Screen 18: **End**

Thank you very much for participating in this experiment today. Please fill out the receipt at your desk, including your total income. Please then remain seated for a moment. We will ask you to come to the adjoining room according to the order of your cubicle numbers. There, you will receive your payment.

Total income: t.

E.1.2 Passive players

Screen 1:
See screen 1 of Active players (see above).

Part 1: Lying-Dictator Game
Screen 2: **Instructions Part One**
In part one of the experiment, most participants in this experiment take part in a task. However, you and three other participants in this session have been randomly chosen **not** to take part in this task. Instead, your earnings will be determined in part two and three.

However, on the next screen we will nevertheless present you the instructions for the task that the other participants are completing in part one. You should read these instructions, but keep in mind that these instructions are for your information only. You will not be asked to take part in the task.

Screen 3: **Please read the instructions. You will, however, not make a decision.**
See screen 3 of Active players (see above).

Screen 4-7:
No participation.

Part 2: Spectator Lying-Dictator Game
Screen 8: **Instructions Part Two**
In part two of the experiment, you are assigned to the role of Participant A/B. In this part you will be matched with one other participant: Participant B/A. Like you, Participant B/A did not make a decision in part one.

In this part of the experiment, your earnings will depend on the decision made by one of the participants who took part in part 1 and now will make a decision in part 2 of the experiment. We will present their instructions for part 2 of the experiment on the next screen. However, keep in mind that you once again will **not** make a decision yourself.

Screen 9: **Please read the instructions. You will, however, once again not make a decision.**
See screen 9 of Active players (see above).

Screen 10-12:
No participation.

Part 3: Norm elicitation.

Screen 13: **Instructions Part Three**

In this part of the experiment **you make a decision** yourself.

See screen 13 of Active players for the rest of the screen.

Screen 14:

See screen 14 of Active players (see above).

Screen 15: **Your income today**

Thank you! For completing the experiment you receive an additional payment of 2 €. Your income today hence consists of the following:

Show-up fee: 5 €

In part 2 of the experiment the participant reported: "The number I saw was z." You were Participant A/B. Hence you will receive z/(10-z) and the Participant B/A will receive (10-z)/z.

In part 3 of the experiment option q was randomly chosen for payment. You indicated that you believed this report to be "p." This was (not) the rating chosen by most other participants. Hence you earn s.

Payment for completing the experiment: 2 €.

Your total payment is therefore: t.

Please answer some questions on the next screen, before you receive your payment.

Questionnaire

Screen 16:

No participation.

Screen 17-18:

See screen 17-18 of Active players (see above).

E.2 Instructions LYING GAME and DICTATOR GAME

Here we present the instructions for the LYING GAME and the DICTATOR GAME. The sequence of Part 1 and Part 2 was randomized.

Screen 1: **General instructions**

Welcome to today's experiment

For showing up today, you will receive a 5 € show-up fee. In addition to this amount, you can earn money during the experiment. The amount of money you earn will depend on the decisions that you and/or other participants make.

During this experiment you are not allowed to use electronic devices, or to communicate with other participants. Please only use the programs provided for the experiment. If you have a question, please remain seated and raise your hand. We will then come to you and will answer your question in private. Please do not ask your question out loud. If you violate these rules you will be excluded from the experiment and the payment.

As with all experiments in this experimental laboratory, there is no deception: That means that whenever we tell you that something will happen if you make a certain choice, then that thing will happen if you make that choice.

This experiment consists of two parts. You will receive the instructions for the first part on the next screen. The instructions to part two will be provided after you have finished part one. For simplicity, we only use the male terminology; of course any explanation is also meant to include women.

Please note that at the end of the experiment either part 1 or part 2 of the experiment will be payoff-relevant and determine your payment. There is a 50% chance that part 1 will be payoff-relvant and determines your payment. There is a 50% chance that part 2 will be payoff-relvant and determines your payment. At the end of the experiment, you will be informed about which of the two parts determines your payment.

Part 1

Screen 2: **Instructions Part One**

In part one of the experiment you will see 11 boxes with numbers hidden in them. The numbers in the boxes are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 and they are placed in a random order. We will ask you to click on one box. Once you click on the box, you will see a number that we ask you to remember and later report to us. The number you report to us determines how much money you will be paid.

The payment corresponding to each number is as follows:

Number Reported	You receive
0	0€
1	1€
2	2€
3	3€
4	4€
5	5€
6	6€
7	7€
8	8€
9	9€
10	10€

After clicking on a box and seeing a number, you can make your report in one of four ways:[28]

- Report: "The number I saw was [number seen]."
- Report: "The number I saw was 5."
- Report: "The number I saw was 10."
- Report: "The number I saw was 0."

Note that this means that potentially two of the reports are the same. If this part of the experiment will be selected, your report ("The number I saw was ...") will determine your payment.

If you have any questions please raise your hand. We will come to you and answer them in private.

Screen 3:

Please click on one of the boxes:

Screen 4:

Please choose the report you would like to make. On the right you see the payment implied by each report.

[28]The order of the first two options was randomized and then kept constant throughout the instructions.

Your report	You receive
○ Report: The number I saw was d.	d €
○ Report: The number I saw was 5.	5 €
○ Report: The number I saw was 10.	10 €
○ Report: The number I saw was 0.	0 €

Screen 5: **End Part One**

The first part of the experiment has finished. The second part will begin shortly. Please press "Continue."

Part 2

Screen 6: **Instructions Part Two**

In part two of the experiment, you can choose between four different options. The option you choose determines how much money you and another participant will be paid.

The four options are as follows:

Option	You receive	Other participant receives
1	d €	(10-d) €
2	5 €	1 €
3	10 €	0 €
4	0 €	10 €

If d equals 0, 5 or 10: Please not that this means that two of the options are the same.

The other participant will also simultaneously complete the same task. If this part of the experiment is selected, either your decision or the other participant's decision will be implemented. There is a 50% chance that your decision will be implemented. In this case your payment is determined by your own decision only. There is a 50% chance that the other participant's decision will be implemented. In this case your payment is determined only by the other participant's decision. At the end of the experiment you will be informed whose decision is implemented.

If you have any questions please raise your hand. We will come to you and answer them in private.

Screen 7:

No participation.

Screen 8:

Please choose one of the four options.

Option	You receive	Other participant receives
◦ 1	d €	(10-d) €
◦ 2	5 €	1 €
◦ 3	10 €	0 €
◦ 4	0 €	10 €

Screen 9: **Your income today**

Thank you! For completing the experiment you receive an additional payment of 2 €. Your income today hence consists of the following:

Show-up fee: 5 €

If randomly selected part is the LYING GAME*:* The computer randomly selected that part 1 determines your payment. In part 1 you reported: "The number I saw was x." Hence you earn x.

If randomly selected part is the DICTATOR GAME *and the participant is selected as decision maker:* The computer randomly selected that part 2 determines your payment. The computer randomly selected your decision to determine the payment. Hence you earn x.

If randomly selected part is the DICTATOR GAME *and the participant is not selected as decision maker:* The computer randomly selected that part 2 determines your payment. The computer randomly selected the other participant's decision to determine the payment. Hence you earn x.

Payment for completing the experiment: 2 €.

Your total payment is therefore: t.

Please answer some questions on the next screen, before you receive your payment.

Questionnaire

Screen 10: **Questionnaire**

Please indicate your gender.

How old are you (in years)?

What do you study? In case you do not study, please enter your current occupation.

In which term are you? In case you do not study, please enter a 0.

How often have you previously participated in an experiment in this laboratory? In case you have never participated, please enter a 0.

How often have you previously participated in a similar experiment? By similar experiment we mean for instance an experiment in which you had to roll a die and report the outcome of the die roll or something similar.

Screen 11: **End**

Thank you very much for participating in this experiment today. Please fill out the receipt at your desk, including your total income. Please then remain seated for a moment. We will ask you to come to the adjoining room according to the order of your cubicle numbers. There, you will receive your payment.

Total income: t.