WZB
Wissenschaftszentrum Berlin
für Sozialforschung

Kai Barron
Robert Stüber
Roel van Veldhuizen

Navigating Moral Trade-Offs

Discussion Paper

SP II 2024–301

November 2024

WZB Berlin Social Science Center

Research Area
Markets and Choice

Research Unit
Economics of Change

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Kai Barron, Robert Stüber, and Roel van Veldhuizen
Navigating Moral Trade-Offs

Affiliation of the authors:

Kai Barron
WZB Berlin Social Science Center

Robert Stüber
University of Warwick

Roel van Veldhuizen
Lund University

Wissenschaftszentrum Berlin für Sozialforschung gGmbH
Reichpietschufer 50
10785 Berlin
Germany
www.wzb.eu

Abstract

Navigating Moral Trade-Offs

by Kai Barron, Robert Stüber, and Roel van Veldhuizen*

An extensive literature documents that people are willing to sacrifice personal material gain to adhere to a moral motive. However, less is known about the psychological mechanisms that operate when two moral motives come into conflict. We hypothesize that individuals adhere to the moral motive that aligns with their self-interest. We test this hypothesis using experiments that induce a conflict between two of the most-studied moral motives: fairness and truth-telling. Consistent with our hypothesis and across experiments, our results show that individuals do prefer to adhere to the moral motive that is more aligned with their self-interest.

Keywords: Moral dilemmas, Dictator game, Lying game, Motives, Motivated reasoning

JEL classification C91; D01; D63; D90

---

*E-mail: kai.barron@wzb.eu, robert.stueber@wbs.ac.uk, roel.van_veldhuizen@nek.lu.se. 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 Kübler, Max Lobeck, George Loewenstein, Peter Schwardmann, Ivan Soraperra, Bertil Tungodden, and Georg Weizsäcker for valuable comments as well as Holly Baker for copy editing. 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. Roel van Veldhuizen thanks the Jan Wallander and Tom Hedelius Foundation for generous financial support. The two experiments in this study were preregistered at the AEA registry astrials no. AEARCTR-0003617 (Laboratory Experiment) and AEARCTR-0013827 (Online Experiment) respectively. This research was conducted under approval from the WZB Berlin Ethics Review Board. This paper was previously circulated under the title "Moral Motive Selection in the Lying-Dictator Game."

1 Introduction

There is ample evidence that many individuals are willing to sacrifice personal gain to adhere to a moral principle or motive. Individuals buy fair-trade goods, donate to charity, and refuse to engage in profitable but unethical behavior such as tax evasion. To model and quantify people's 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 to reduce inequality (Fehr and Schmidt, 1999) or to avoid lying (Fischbacher and Föllmi-Heusi, 2013).

However, individuals frequently encounter situations where multiple moral principles compete, leading to conflicting ethical imperatives. For example, consumers might struggle to choose between purchasing a fair-trade product or an organic one. Managers may face dilemmas between improving the working conditions of their employees and ensuring job security in the company. Politicians often grapple with decisions that pit different fairness principles—such as equality of outcomes versus equality of opportunities—against each other. In such ethical dilemmas, individuals may tend to prioritize the moral motive that aligns most closely with their personal interests. When choosing between different forms of ethical consumption, for example, consumers may favor the one with the lowest price. Managers may focus on objectives that increase their own bonus payments. When deciding between different policy positions, politicians may adhere to the ethical principle that increases their popularity with the current electorate.

We study such behavior in experiments that involve a conflict between two of the most well-studied moral motives: truth-telling and fairness. We design a setting 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's (2013) lying (or cheating) game. As in the lying game, decision makers observe a random number draw and are asked to report its outcome. 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. While truthfully reporting the observed draw may satisfy a desire to be honest, it may also lead to an unfair (unequal) allocation. Thus, this LYING-DICTATOR GAME generates a conflict between truth-telling and equality motives. Our identification strategy relies on the random draw, which creates exogenous variation in the financial cost of adhering to the truth-telling motive. For low random draws, it is more costly to tell the truth, and an individual who wants to behave in a moral way but cannot simultaneously satisfy both moral motives may adhere to the fairness motive. By contrast, for high random draws, adhering to the fairness motive is more costly, and hence the individuals may choose to tell the truth instead.

We first conduct a laboratory experiment with German students. Our results reveal strong evidence that individuals adhere to the less-costly moral motive. Among participants with a high random draw, 75 percent choose to tell the truth and only 9 percent choose the equal division. By contrast, among participants with the low random draw, 14 percent choose to tell the truth and 47 percent choose to equalize payoffs. To provide behavioral benchmarks, we run additional sessions in which participants play the standard dictator and lying games. We find that individuals in the LYING-DICTATOR GAME with low random draws behave similarly to individuals playing the classic dictator game, whereas individuals in the LYING-DICTATOR GAME with high random draws behave similarly to individuals in a standard lying game.

We then investigate two potential underlying mechanisms. First, participants may have *stable moral preferences*. According to this hypothesis, the random draw affects the opportunity cost of adhering to a particular moral motive but does not affect moral preferences per se. Second, participants may engage in *motivated reasoning*. According to this perspective, the random draw directly affects moral preferences by inducing participants to change their utility function to put greater weight on the moral motive that is cheaper for them to satisfy.

We provide evidence towards distinguishing between these mechanisms in two ways. First, after completing the (first-party) LYING-DICTATOR GAME, participants take part in a spectator (third-party) version of the same game where their reports determine the payments of two other individuals. If moral motive selection is driven by motivated reasoning, the random draw in the LYING-DICTATOR GAME should affect the weights attached to different moral motives in their utility function. If these adjusted utility weights persist over time and translate to the third-party version of the same game, then they should spill over to the SPECTATOR LYING-DICTATOR GAME. However, 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. Second, we study moral motive selection using a theoretical model of moral decision-making and compare a version of the model with stable preferences to a version with motivated reasoning. Whereas both versions of the model generate a moral motive selection effect of similar size, the stable preferences model is better able to accommodate the reduced frequency of selfish choices we observe in the LYING-DICTATOR GAME compared to the two baseline games.

We then conduct a second experiment online with a large general-population sample from the U.S. We increase the scope for spillovers in the spectator game by letting participants choose their report *before* they observe the random draw and varying whether the second (spillover) game is a spectator game (as in the lab experiment) or another first-party LYING-DICTATOR GAME. Second, we vary whether participants are restricted to the numbers corresponding to specific moral motives (as in the lab experiment) or can pick any number, allowing us to assess whether our results are replicated in a game that is closer to the way the dictator and lying game are typically implemented. Consistent with the results from the lab, we find strong evidence of moral motive selection and no evidence of spillovers in all treatments. Hence, our results appear to be robust to various design choices, sampling procedures, and the choice of participant pool.

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).[1] More recent work examines distributional preferences in settings where income is earned (e.g., Frohlich et al., 2004; Cappelen et al., 2007; Almås et al., 2020). In contrast to our study, this body of work has identified potential conflicts between motives in one domain (fairness).

Evidence regarding a *truth-telling motive* 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 many people are willing to significantly reduce their earnings in order to avoid telling a lie (see Abeler et al., 2019 and Gerlach et al., 2019 for recent reviews). 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. Whereas existing studies find little effect of introducing negative externalities of dishonesty to others in the lying game (Fischbacher and Föllmi-Heusi, 2013; Gneezy and Kajackaite, 2020), we document a strong impact if the trade-off between fairness and lying is made explicit.[2]

Our results also contribute to the literature on motivated reasoning and motivated beliefs. These studies show that an individual's self-interest may distort her beliefs (e.g., Di Tella et al., 2015; Palma and Xu, 2019; Gneezy et al., 2020; Schwardmann et al., 2022; Saccardo and Serra-Garcia, 2023; Gino et al., 2016), affect the way that she gathers information (e.g., Babcock et al., 1996; Dana et al., 2007; Ambuehl, 2021; Exley and Kessler, 2023) and bias her judgment of different *fairness* ideals (e.g., Messick and Sentis, 1979; Babcock et al., 1995; Konow, 2000; Rodriguez-Lara and Moreno-Garrido, 2012; Capraro and Rand, 2018; Gneezy et al., 2019; Amasino et al., 2021; Neuber, 2021). We extend this line of work by examining whether people distort the importance of moral motives from different moral domains.

[1]This inspired models of warm-glow giving (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).

[2]For the deception game, studies have typically found externalities to matter (e.g.,Gneezy, 2005; Erat and Gneezy, 2012; Alempaki et al., 2018; Michailidou and Rotondi, 2019).

2 Laboratory Experiment

2.1 Experimental Design

2.1.1 Experimental Games

Our main interest lies in the Lying-Dictator Game in which a decision maker (DM) is asked to report a random number that determines her and another participant's payment. We compare the decision-making in the Lying-Dictator Game to the Dictator Game and a version of Fischbacher and Föllmi-Heusi's (2013) Lying Game. Finally, we also consider behavior in a Spectator Lying-Dictator Game. This game is equivalent to the Lying-Dictator Game but the number the DM reports now affects the payment of two other players.

We implemented these four games in two separate between-subject treatment conditions. In our main 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 treatment allows us to compare the results of the Lying-Dictator Game to the baseline games studied in previous work.

2.1.2 Treatment Lying-Dictator

In treatment Lying-Dictator, participants were assigned to be either Active players or Passive players. Active players served as a DM in both parts. Passive players did not participate in part 1 and served as the third party (player A or player B) in part 2.

Part 1: The Lying-Dictator Game
At the start of the experiment, Active players received the instructions for the Lying-Dictator Game. Specifically, they were informed that they would be presented with a screen containing 11 boxes, as shown in 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 to Gneezy et al.'s (2018) implementation of the LYING GAME. Note that a key advantage of this design is that it allows us to record both the report r and the value of the random draw d. Participants could then report the number in one of four ways:

1. Tell the truth and report: "The number I saw was [*number seen*]."[3]

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 moral motives salient to participants. Second, it reduces the number of available reports. This prevents participants from making intermediate choices that do not correspond to following one of the relevant moral motives directly, potentially sharpening our analysis by making it easier to classify responses as either truth-telling, equalizing, or payoff-maximizing. In our online experiment we show that restricting the action space does not impact our results in a meaningful way (see Section 3 below).

To maximize the number of participants in the role of the DM, we used the strategy method (role uncertainty). We asked all Active players to make a decision as if they were the DM but told them that there was only a 50% chance that their choice would be implemented; otherwise, they would act as recipients for another DM's decision. 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.

Part 2: The Spectator Lying-Dictator Game

The 20 Active players in each session 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 to then report this number in one of four ways:

1. Tell the truth and report: "The number I saw was [*number seen*]."

[3]The order of the first two reports (truth-telling or equalizing payments) was randomized between participants.

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.”

The main difference to part 1 is that the number reported does not affect the DM’s monetary payoff but those of two other (passive) players. Therefore, the DM’s decision in part 2 is a pure choice between moral motives, without her monetary self-interest at play. At the end of each session, two of the Active players were randomly chosen and their decisions determined the payment for the two pairs of Passive players. As in part 1, the reports that were implemented were also sent to the recipients.

Part 3: The Norm Elicitation Task

All participants then received the instructions to part 3, the NORM ELICITATION TASK. Following Krupka and Weber (2013), we asked participants to consider four possible reports made by a hypothetical participant who faced the SPECTATOR LYING-DICTATOR GAME and received a random draw of 8 (which generates a clear conflict between truth-telling and equality). We then asked participants to rate each of the four possible reports in terms of its “social appropriateness.” on a 6-point scale. 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.

2.1.3 Treatment BASELINE

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. In contrast, the report no longer affects another participant's payment, which removes the equality motive.

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. We asked participants to choose between four allocations corresponding to equality, a random draw, payment maximization, and payment minimization, respectively. As in the LYING-DICTATOR treatment, we asked participants to make decisions as if they were the DM, but we told them that there was only a 50% chance that their decision would actually be implemented, and otherwise they would act as recipients. The key difference is that the random draw was done behind the scenes by the computer, instead of being done explicitly by the participant. This procedure removes truth-telling as a moral motive.

2.1.4 Remaining Procedures, Questionnaire, and Payment

All participants subsequently went through a post-experimental questionnaire. Sessions in treatment LYING-DICTATOR took approximately 30 minutes. The average payment was EUR 12.93. Payments ranged from EUR 7 to EUR 19. Sessions in treatment BASELINE took approximately 20 minutes. The average payment was EUR 13.37, and payments ranged from EUR 7 to EUR 17.

We conducted our lab experiment at the WZB-TU laboratory for experimental economics in Berlin. We programmed the experiment using zTree (Fischbacher, 2007) and invited participants using ORSEE (Greiner, 2015). Instructions were provided on-screen and are reproduced in Online Appendix C1.

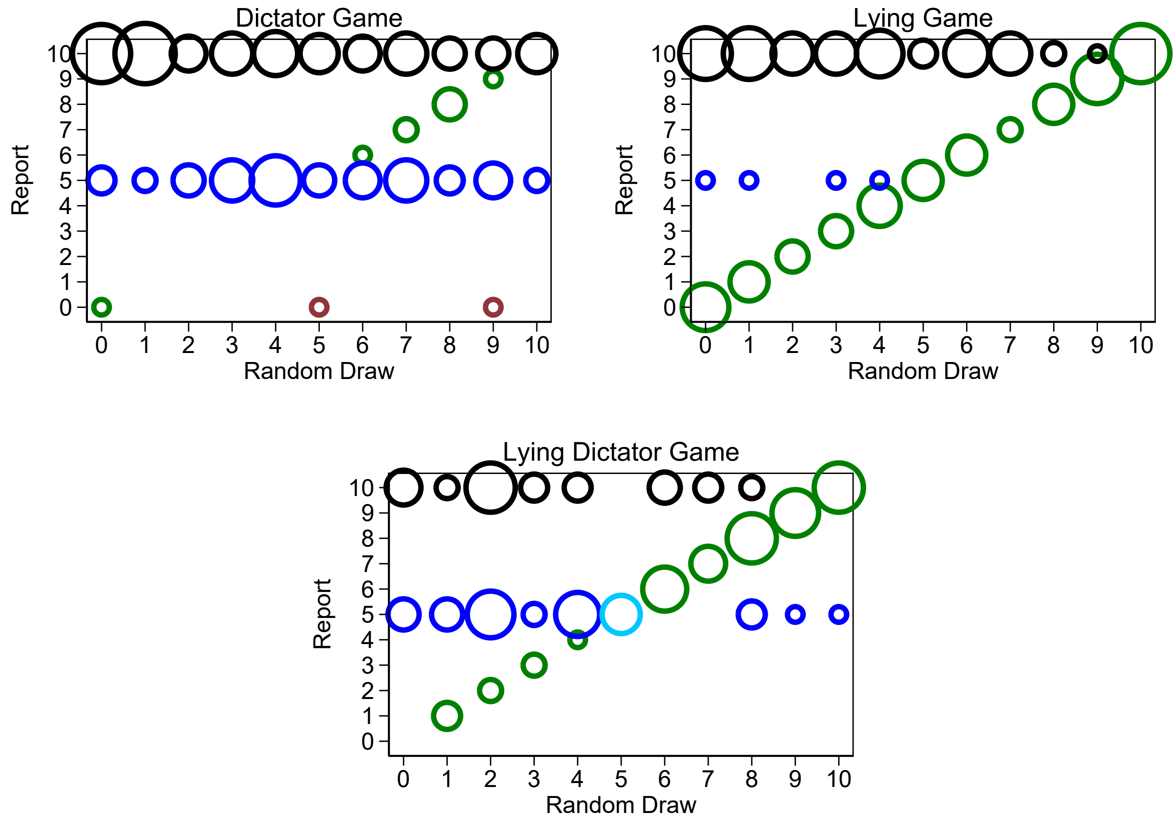
We preregistered the experimental design of the LYING-DICTATOR treatment at the AEA registry, including a pre-analysis plan, power calculation, and a detailed discussion of the relationship to the two classical benchmark games. We reproduce the pre-registration in Online Appendix D1.

2.2 Results

2.2.1 Moral Motive Selection

Figure 2 provides a visual representation of our data (we present descriptive statistics in Appendix A). 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. The behavior observed in the two games is similar to that documented in previous work (see Online Appendix A for details). In the DICTATOR GAME, a large share of participants chooses to either equalize payments (a report of 5) or to maximize their own payoff (a report of 10). In the observed LYING GAME, the majority of participants choose either to tell the truth (the diagonal entries) or to maximize their own payoff (a report of 10).

Figure 2: Distributions of choices across the three games



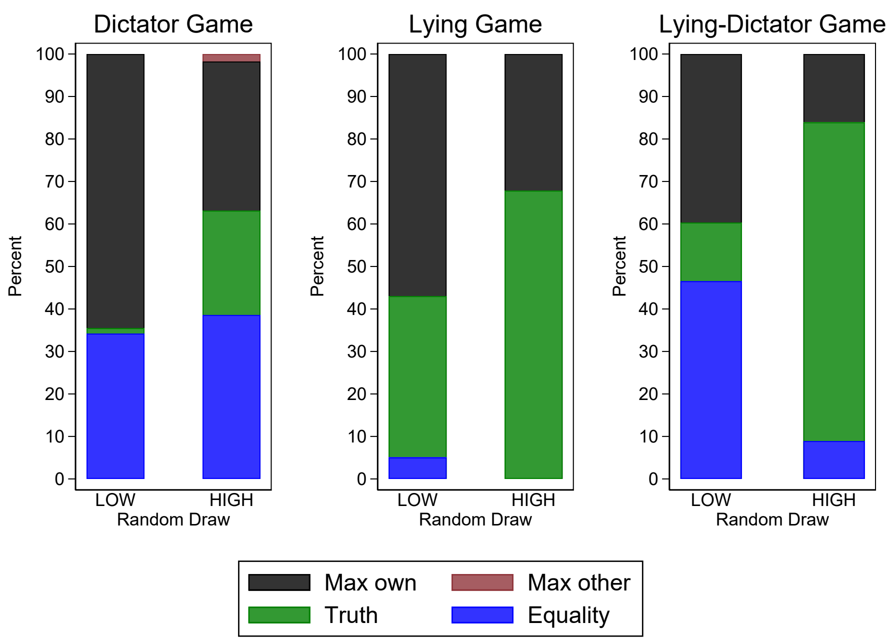
Note: The figure shows the distribution of reports by random draw for each of the three games. Circle sizes correspond to the number of participants, colors to the different motives.

The lower panel of Figure 2 presents the data of the LYING-DICTATOR GAME. For random draws lower than 5 (LOW draws), almost half (47%) of our participants select the motive that most closely aligns with their self-interest (equality), and only 14% choose to tell the truth (diff.: 32.76 pp, two-tailed test of proportions, $p < 0.001$). By contrast, for random draws greater than 5 (HIGH draws), the vast majority (75%) tell the truth and very few (9%) choose to equalize payoffs (diff.:

66.07 pp, two-tailed test of proportions, $p < 0.001$). These results imply that moving from a LOW to a HIGH random draw significantly increases the rate of truth-telling (diff.: 61.21 pp, two-tailed test of proportions, $p < 0.001$) and significantly decreases the propensity to choose equality (diff.: 37.62 pp, two-tailed test of proportions, $p < 0.001$). Therefore, participants predominantly select the moral motive that most closely aligns with their self-interest.

We can also check whether the degree to which behavior in the LYING-DICTATOR GAME resembles the two baseline games depends on the random draw. If participants adhere to the motive that most closely aligns with their self-interest, behavior after a LOW draw (where equality is least costly) should resemble the DICTATOR-GAME (where equality is the only moral motive). In contrast, behavior after a HIGH draw (where honesty is least costly) should resemble the LYING-GAME (where honesty is the only moral motive). Consistent with this hypothesis, Figure 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.

Figure 3: Moral motive choices across game types



Notes: The figure shows the distribution of moral motive choices for a LOW (< 5) draw or a HIGH (> 5) 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.

Taken together, both effects imply that the change in moral motive choices between a LOW and a HIGH draw is significantly more pronounced in the LYING-DICTATOR GAME than in the two baseline games. At the same time, the figure also illustrates that selfish behavior is less common in the LYING-DICTATOR GAME than in the two baseline games (27% in the LYING-DICTATOR GAME vs 52% in the DICTATOR GAME, $p < 0.001$, and 27% vs 46% in the LYING GAME, $p < 0.002$). We discuss these latter two findings in more detail in Online Appendix B.

2.2.2 Mechanisms of Moral Motive Selection

These results demonstrate that many participants choose to adhere to the moral motive that is least costly for them to satisfy. In this section, we investigate the 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 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 affects 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.

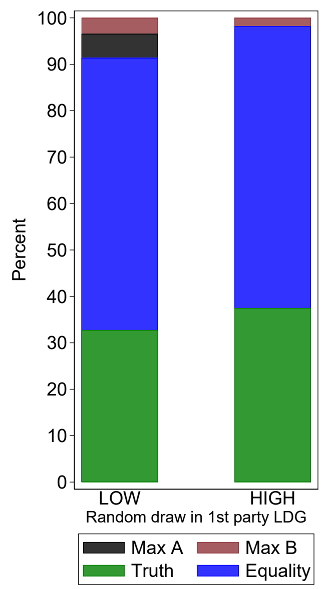
The Spectator Lying-Dictator Game

First, we use the SPECTATOR LYING-DICTATOR GAME to distinguish between these two mechanisms experimentally. 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. Under the assumption that these adjusted weights on the motives persist into the subsequent spectator game, these participants should then also be more likely to choose 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 effects.

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. The key result is that the random draw in the LYING-DICTATOR GAME does not have a statistically significant impact on the moral 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, $p = 0.820$).[4] 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 moral motives (two-tailed Mann-Whitney U test, $p = 0.508$). Therefore, we observe no evidence that the moral motive selection we observe in the first-party LYING-DICTATOR GAME spills over into subsequent choices and norm perceptions in the SPECTATOR LYING-DICTATOR GAME in a manner that would be predicted by motivated reasoning. However, the latter prediction requires assuming that the adjusted weights on the motives persist over time and carry over into the subsequent game.

Figure 4: Moral motive choices in the SPECTATOR LYING-DICTATOR GAME



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

A Model of Moral Decision-Making

To further distinguish between and formalize these two explanations, 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

[4]The results are robust to excluding participants who made selfish choices in part 1 or to only focusing on those participants, and to controlling for the exact random draw. Participants who tell the truth in the LYING-DICTATOR GAME 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, $p = 0.672$).

social preferences, respectively. We incorporate motivated reasoning by allowing participants to adjust the relative weight given to each moral motive. The details of the model are in Appendix B.

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. 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 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 moral 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 participants adhering to the cheaper moral motive 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, in line with the absence of spillovers observed in the SPECTATOR LYING-DICTATOR GAME, the evidence is therefore more consistent with stable preferences than with the motivated reasoning explanation.

3 Online Experiment

3.1 Design

Two features of our lab experiment deserve closer investigation. First, the action space in the LYING-DICTATOR GAME is restricted: participants cannot choose any number but have to choose a report corresponding to one of four motives. This makes moral motives very salient. This may prompt participants to explicitly consider the trade-offs between motives and respond to the cost of each motive in a way that may not occur if the action space is not truncated. Second, in the SPECTATOR LYING DICTATOR GAME, the DM faces a new random draw and her own payoff is no longer at stake. These design elements may reduce the similarity in moral context between SPECTATOR LYING DICTATOR GAME and LYING DICTATOR GAME, and thereby limit the scope for spillovers.

We therefore conducted an experiment to assess the robustness of our results with respect to these design features. The design of the experiment closely follows the design of the first experiment. The three main differences are the following. First, the experiment was conducted online with a large U.S. sample. Second, in the second-stage game, we ask participants to make their decision before observing the draw. Third, we run a 2x2 design that varies the action space (restricted or unrestricted) and the second game (SPECTATOR LYING-DICTATOR GAME or first-party LYING-DICTATOR GAME). Further design details are presented in Appendix C.

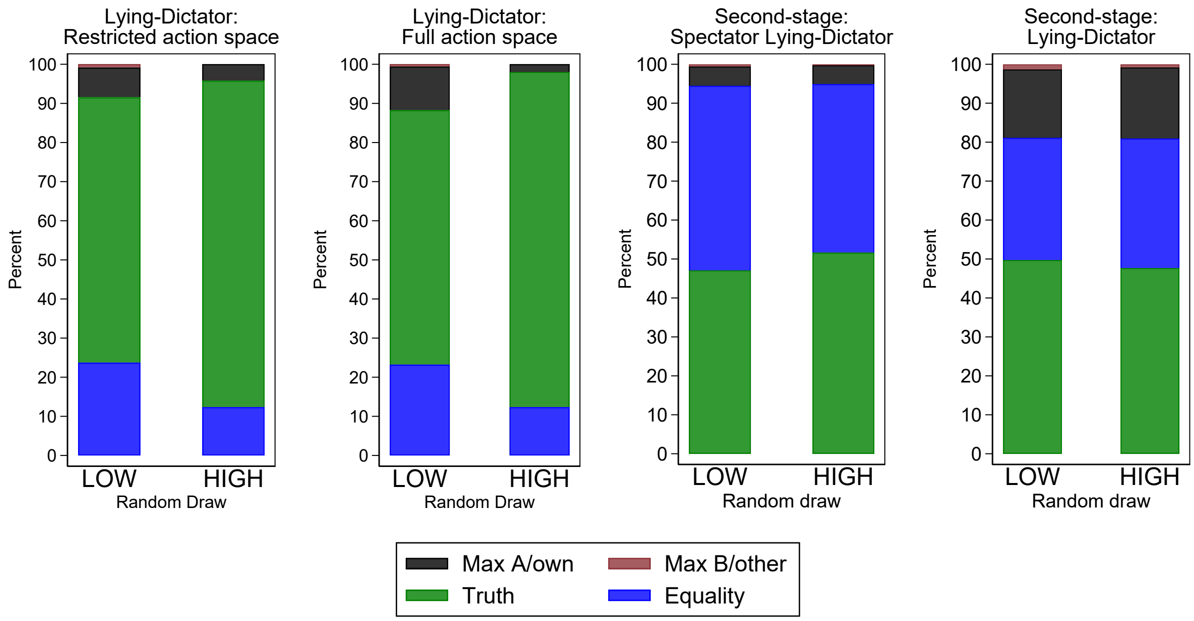
3.2 Results

Replicating our laboratory results, we find higher truth-telling rates and lower rates of equality choices after HIGH draws than after LOW draws in the LYING-DICTATOR GAME (see Figure 5). In the two treatments with the restricted action space, which most closely resemble our laboratory experiment, participants with HIGH draws are 17 percentage points more likely to tell the truth ($p < 0.001$) and 10 percentage points less likely to choose equality ($p < 0.001$). In the two treatments with the full (unrestricted) action space we find a difference of 22 percentage points ($p < 0.001$) in truth-telling rates and 10 percentage points ($p < 0.001$) in equality, respectively. Overall, we therefore also find evidence of moral motive selection when we (i) use an online sample and (ii) provide participants with the full action space.

We also again find no evidence that the random draw from the LYING-DICTATOR GAME affects behavior in the second game. As Figure 5 shows, the random draw in the LYING-DICTATOR GAME does not have any effect on the moral motive selected in the second-stage game. The share of truth-telling and equality choices are almost identical for LOW and HIGH random draws in the first-stage LYING DICTATOR GAME. This holds true both when the second-stage game consists of a SPECTATOR LYING DICTATOR GAME and a LYING DICTATOR GAME. We present a more detailed analysis of both motive selection and spillovers in Appendix C.3.

Overall, the results of the online experiment therefore provide further support that people select the cheaper moral motive and that their choices do not spill over to subsequent choices. Hence, we replicate the main results from the laboratory experiment despite studying a different population and changing several features of our design. Finally, the high truth-telling rates for low random draws are more consistent with the stable moral preference explanation of our results, as we discuss further in Appendix B.

Figure 5: Moral motive choices across treatments of the Online Experiment



Note: The figure shows the distribution of moral motive choices for participants who drew a LOW (< 5) or HIGH (> 5) number in the first-stage LYING-DICTATOR GAME by treatment. Under the full action space, the participants who choose a report that does not align with any of the motives (5%) are excluded.

4 Conclusion

How do people decide 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 complying with the moral motive that is cheapest to satisfy. We test for such moral motive selection using a simple game that is isomorphic to the classic dictator game and standard lying game. The key 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.

In both a laboratory and an online experiment, we find that participants tend to comply with the more convenient moral motive. We further show 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, our experimental design allows us to investigate the mechanisms behind moral motive selection in two ways. The result, from experimental treatment variation in the lab and online experiments as well as from models that contrast stable moral preferences and motivated reasoning, point towards a limited role of motivated reasoning in this context.

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. 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). This induces socially desirable truth-telling. 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. 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 in our own LYING GAME and than commonly found in the literature.

The consequences of this type of behavior 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. Through this lens, the observed "moral motive selection" can be thought of as being analogous to a consumption decision, where individuals choose to "buy" compliance with the cheaper moral motive. A worrying consequence is that many individuals are likely to choose the cheapest moral action in their choice set. This means that in multiple moral motive settings, we will see far less adherence to "expensive" moral actions than in single moral motive settings.

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.

ALEMPAKI, D., G. DOĞAN, AND S. SACCARDO (2018): "Deception and reciprocity," *Experimental Economics (forthcoming)*.

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, 464–477.

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.

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.

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.

DANA, J., R. A. WEBER, AND J. X. KUANG (2007): "Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness," *Economic Theory*, 33, 67–80.

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.

ENGELMANN, D. AND M. STROBEL (2004): "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments," *American Economic Review*, 94, 857–869.

ERAT, S. AND U. GNEEZY (2012): "White lies," *Management Science*, 58, 723–733.

EXLEY, C. L. AND J. B. KESSLER (2023): "Information Avoidance and Image Concerns," *The Economic Journal*, 133, 3153–3168.

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.

FROHLICH, N., J. OPPENHEIMER, AND A. KURKI (2004): "Modeling other-regarding preferences and an experimental test," *Public Choice*, 119, 91–117.

GERLACH, P., K. TEODORESCU, AND R. HERTWIG (2019): "The truth about lies: A meta-analysis on dishonest behavior." *Psychological Bulletin*, 145, 1–44.

GINO, F., M. I. NORTON, AND R. A. WEBER (2016): "Motivated Bayesians: Feeling Moral While Acting Egoistically," *Journal of Economic Perspectives*, 30, 189–212.

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

GNEEZY, U. AND A. KAJACKAITE (2020): "Externalities, stakes, and lying," *Journal of Economic Behavior & Organization*, 178, 629–643.

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–S300.

KONOW, J. (2000): "Fair shares: Accountability and cognitive dissonance in allocation decisions," *American Economic Review*, 90, 1072–1091.

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.

MICHAILIDOU, G. AND V. ROTONDI (2019): “I’d lie for you,” *European Economic Review*, 118, 181–192.

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

PALMA, M. AND Z. P. XU (2019): “Shadow of a doubt: Moral excuse in charitable giving,” *Review of Behavioral Economics*, 6, 133–146.

RABIN, M. (1993): “Incorporating fairness into game theory and economics,” *American Economic Review*, 83, 1281–1302.

——— (1995): “Moral Preferences, Moral Constraints, and Self-Serving Biases,” *Working Paper*.

RODRIGUEZ-LARA, I. AND L. MORENO-GARRIDO (2012): “Self-interest and fairness: self-serving choices of justice principles,” *Experimental Economics*, 15, 158–175.

SACCARDO, S. AND M. SERRA-GARCIA (2023): “Enabling or limiting cognitive flexibility? Evidence of demand for moral commitment,” *American Economic Review*, 113, 396–429.

SCHWARDMANN, P., E. TRIPODI, AND J. J. VAN DER WEELE (2022): “Self-persuasion: Evidence from field experiments at international debating competitions,” *American Economic Review*, 112, 1118–1146.

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.

APPENDIX

A Descriptive Statistics for the Laboratory Experiment

A total of 288 participants took part in the laboratory experiment spread evenly across the two treatments. For the LYING-DICTATOR treatment, we report the descriptive statistics of the 120 Active players. We include all 144 participants of the BASELINE treatment (Table A.1). Nearly all participants were students with an average age of 22.7. 63% of participants were men, and they were approximately in their fifth semester. The most common majors were engineering and related fields (46%), natural sciences (including math and informatics, 21%) and economics or business (17%). Table A.1 shows that the two treatments are well-balanced. Only experience with lying experiments is slightly greater in the Baseline treatment, but only at the 10% level. Likewise, Panel B in Table A.1 shows that the distribution of random draws did not differ significantly across the various treatments and games.

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)	
Engineering	0.462	0.417	0.500	0.176
	(0.500)	(0.495)	(0.502)	
Science	0.205	0.225	0.188	0.452
	(0.404)	(0.419)	(0.392)	
Economics	0.174	0.167	0.181	0.767
	(0.380)	(0.374)	(0.386)	
Other Major	0.159	0.192	0.132	0.186
	(0.366)	(0.385)	(0.340)	
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 (column 5). For the LYING-DICTATOR GAME, only Active players are included.

B Modeling Mechanisms of Motive Selection

The main result of our experiments is that many participants engage in what we call "moral 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.

B.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. Whereas 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's (1999) model of inequity aversion for the equality component and Abeler et al.'s (2019) model of preferences for truth-telling for the truth-telling component. Agent i's utility function is:

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

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 probability of being seen as a liar.[5] Note that we assume that the two moral components are additively separable for simplicity.

Whereas 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 moral motives. In other words, the model assumes that the agent can use motivated reasoning to emphasize the moral motive that agrees with an action and downplay the moral motive that disagree with it. Note that the model holds the total weight awarded to both moral 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 the laboratory experiment. 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:

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

[5] The preferred model in Abeler et al. (2019) 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.

B.2 Calibrating the Model

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 $\alpha_1 \sim U[3, 15]$.[6] Because 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 A.2 in the Online Appendix).[7]

In what follows we will use these calibrated parameters to find the version of the model that can best accommodate the following three stylized facts from the laboratory experiment. (1) Moral motive selection: the frequency of truth-telling (equality) choices in the LYING-DICTATOR GAME is greater for HIGH (LOW) random draws. (2) The frequency of selfish choices is smaller in the LYING-DICTATOR GAME than in the baseline games. (3) There are no spillovers from the random draw in the LYING-DICTATOR GAME to the choices made in the SPECTATOR LYING-DICTATOR GAME. We already discussed in the main text that the stable preference model can best accommodate the third stylized fact, and will therefore focus on the first two facts in what follows below. Similarly, since we calibrate our parameters using laboratory data, we will focus our analysis here on the results of the laboratory experiment. The implications for the online experiment are covered in the discussion section below.

[6] 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 $\theta \sim U[0, 12]$. Because all lies are detected in our experiment, $\Lambda(x_i) = 1$ whenever $x_i \neq \bar{x}_i$. Taken together, this implies that $\alpha_1 = c + \theta \sim U[3, 15]$.

[7] We use the SPECTATOR LYING-DICTATOR GAME for a separate test of motivated reasoning and therefore do not use it to calibrate our model. Calibrating our model using this game would also have required us to make assumptions about how lying costs and inequality aversion parameters translate to a setting where the dictator does not have a personal stake.

B.2.1 Stable Moral Preferences

We can use the 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 are shown in the middle part of Figure B.1. For comparison, we also reproduce our findings from the LYING-DICTATOR GAME (left part of Figure B.1). The figure 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 moral 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). As a result, the overall distribution of choices fits the empirical data from the laboratory experiment quite well.

Nevertheless, the HIGH/LOW effect is somewhat smaller than the effect we find in our data. 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 61 pp, the model only predicts effects of 16 pp and 38 pp, respectively. This suggests that allowing for motivated reasoning may improve the fit of the model by generating a larger moral motive selection effect.

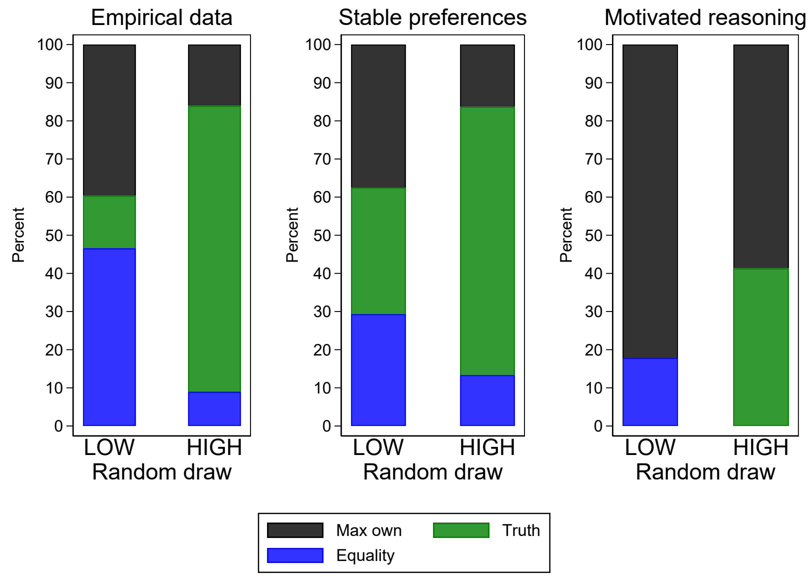
B.2.2 Motivated Reasoning

We allow for motivated reasoning by assuming that participants are able to adjust the relative weight for the moral 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.

Figure B.1: Predictions of models



Note: The left part of the figure ("Empirical data") reproduces our findings for the LYING-DICTATOR GAME (see Figure 3). The middle and right part show 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 and motivated reasoning model. The figure assumes that lying costs and social preferences are independent.

This implies that that the doer is choosing between three options: maximizing her own payoff, equalizing payments, and telling the truth.[8] 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 ($\gamma = 0$). Because 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

[8]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.

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.[9]

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

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

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 (because $(2 - \gamma)\alpha_3 > 0.5$) 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.

The model predictions are shown in the right part of Figure B.1, 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 (16 pp and 39 pp, respectively). That is, allowing for motivated reasoning does not increase the predicted size of the moral 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

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

number of cases. This also implies that the motivated reasoning model predicts a greater frequency of selfish choices than in the two baseline games, which runs counter to what we observe in our data.

B.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 moral motive selection (HIGH/LOW) effect in the LYING DICTATOR GAME. 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 LYING DICTATOR GAME more closely than the model with motivated reasoning. As a result, the stable preference model is better able to accommodate the overall frequencies of observed choices in the data (see Figure B.1). Similarly, the stable preference model can also better accommodate the results from the SPECTATOR LYING-DICTATOR GAME (no spillovers), as discussed in section 2.2.2 in the main text.

The stable preference model can also better accommodate the results from the online experiment. In particular, the motivated reasoning model (with $c(y) = 0$) predicts that no one will ever choose to adhere to the more expensive moral motive. Yet in the online experiment, the *majority* of participants choose to tell the truth after a LOW random draw. By contrast, under stable preferences, the high overall truth-telling rates and relatively small motive selection (HIGH/LOW) effect) in the LYING DICTATOR GAME can be accommodated by a model with a high lying cost and low inequality aversion.

Why does motivated reasoning not increase the moral motive selection (HIGH/LOW) effect in our model? Intuitively, motivated reasoning increases the scope for moral motive selection by allowing participants to adjust γ to increase the appeal of the less costly moral 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 moral 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 moral motive selection (HIGH/LOW) effect overall.

Although 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. When we instead assume that social preferences and lying costs are perfectly rank-correlated, we obtain very similar results as shown in Figure B.2. 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 moral motive selection (HIGH/LOW) effect, but the predicted effect tends to be smaller than the one we observe in the data from the lab. This may be the result of the specific way we calibrated our model. In particular, we can also consider the following simple model:

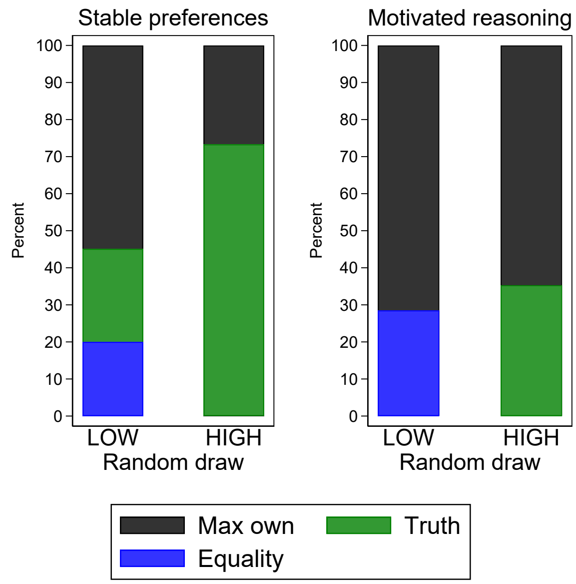
$$U(x_i) = x_i - \gamma[\alpha I(x_i \neq \bar{x}_i)] - (2-\gamma)[\alpha I(x_i \neq 5)] \tag{8}$$

This specification assumes that participants pay a fixed moral cost of α for any moral 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 thirds has $\alpha \sim U[3, 15]$ accommodates the data from the baseline games well.

Figure B.3 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 moral motive that is cheapest for them to satisfy even without motivated reasoning. Motivated reasoning slightly increases the scope for selfish behavior by allowing participants to lower the weight on the more attractive moral motive as a way to increase the relative appeal of the selfish option.[10] In keeping with previous models, however, motivated reasoning has only a mini-

[10]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 with the stable preference case, the planner will inflate the weight put on the more costly moral 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.

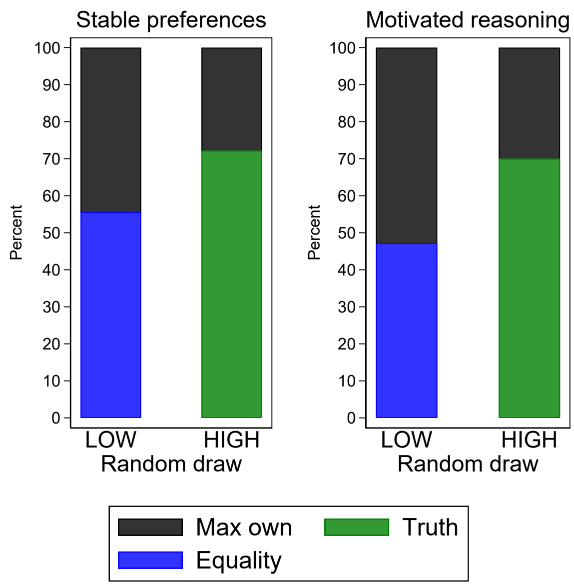
Figure B.2: Predictions of models under correlated preferences



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.

mal effect on the size of the HIGH/LOW effect, slightly reducing it. Overall, the simple alternative model fits the data very well, and further suggests that allowing for motivated reasoning is unlikely to affect the size of the HIGH/LOW effect (moral motive selection) in a major way.[11]

Figure B.3: Alternative model predictions



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.

Finally, it is worth noting that the predictions of the stable preferences and motivated reasoning model also differ in terms of the moral motive selected *conditional on not being selfish*. In particular, under moral motive selection, non-selfish participants will always select the cheaper moral motive (truth-telling for high draws, equality for low draws). Under stable preferences, however, participants may still select the more expensive moral motive. In particular, the main model with independent (perfectly correlated) moral costs predicts that 47% (56%) of non-selfish participants

[11]Moral motive selection can also be captured by a model that sees moral motives as constraints rather than preferences (Rabin, 1995). One example is a model that assumes that the moral cost α can be avoided when at least one moral motive is satisfied. The utility function in the LYING-DICTATOR GAME can then be modeled as $U(x_i) = x_i - \alpha * I(x_i \neq \bar{x}_i) * I(x_i \neq 5)$. This specification would lead to predictions that are similar to the models of stable preferences discussed previously in that it predicts a strong HIGH/LOW effect and a *greater* frequency of selfish choices in the LYING-DICTATOR GAME than in the baseline games. When also allowing for motivated reasoning by letting the planner pick the relevant moral motive for the doer, the predictions of this model would be identical to the model of equation 8.

choose to tell the truth for a LOW draw, and 16% (0%) may choose equality for a HIGH draw. In the laboratory 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 moral 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 moral motive.

All in all, the analysis in this section reveals two main results. First, when calibrated using laboratory data, both versions of the model generate a similar HIGH/LOW (moral motive selection) effect in the LYING-DICTATOR-GAME. Second, allowing for motivated reasoning increases the frequency of selfish behavior. This implies that in the laboratory data 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 with the results of spectator games and online experiment, our experimental data are thus more consistent with a model with stable preferences than one with motivated reasoning.

C Online Experiment

In this section, we describe the design and results of the online experiment in greater detail. The experimental design of the online experiment follows the same basic structure as the laboratory experiment, described in the main text. We run a 2x2 design that varies the action space (restricted or unrestricted/full space) and the second game (spectator or first-party LYING-DICTATOR GAME). This gives us the following treatments:

1.	RESTRICTED-SPECTATOR:	LDG/SLDG (restricted action space)
2.	RESTRICTED-FIRSTPERSON:	LDG/LDG (restricted action space)
3.	FULLSPACE-SPECTATOR:	LDG/SLDG (full action space)
4.	FULLSPACE-FIRSTPERSON:	LDG/LDG (full action space)

C.1 Experimental Design

The RESTRICTED-SPECTATOR treatment follows the same structure as our lab experiment. The three other treatments change the action space, the second game, or both. We first describe the design of the RESTRICTED-SPECTATOR treatment, followed by a discussion of the treatment differences. The instructions for all four treatments are provided in Online Appendix C2.

C.2 RESTRICTED-SPECTATOR

Similar to the laboratory experiment, we have Active players and Passive players. The purpose of the Passive players is to serve as recipients of payments generated by the decisions of Active players. Therefore, we focus on the Active players here.

Part 1: The Lying-Dictator Game

At the start of the experiment, Active players received the instructions for the LYING-DICTATOR GAME. Specifically, they were informed that they would be presented with a screen containing 11 boxes. They 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.

As in the laboratory experiment, participants had four options for reporting the number:

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 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.”

We also told participants that after choosing their reports they would be asked to provide a brief written explanation for why they chose these reports. In contrast to the laboratory experiment, Active players always acted as decision makers. The role of recipients was instead played by Passive players.

Part 2: The Spectator Lying-Dictator Game
As in the laboratory experiment, the Active players then received the instructions for Part 2, the SPECTATOR LYING-DICTATOR GAME. The main difference relative to Part 1 is that the number reported does not affect the Active player’s monetary payoff; instead, it affects the payment of two Passive players: Player A and Player B. Similar to Part 1, participants were told that they would have to report a choice 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.”

The key difference between the online experiment and the laboratory experiment is that we used the strategy method in Part 2. This means that Active players made their Part 2 decisions *before* seeing which random number they would draw. Essentially, they made their choice behind the veil of ignorance. We expected that this change would increase the scope for spillovers from Part 1, since it prevents participants from basing their choice (i.e., selected motive) on the random draw in Part 2.

C.2.1 Treatment Variation

We vary two treatment dimensions using a 2x2 between-subjects design. On the first treatment dimension, we vary the *action space*. We do so because the action space is restricted in our lab experiment. This may raise a concern about whether this design feature may be responsible for some of our results. For instance, the restricted action space may increase the salience of the choices that fully correspond to a moral motive by preventing participants from making intermediate choices that do not correspond to such motives directly.

To alleviate this concern, the online experiment features treatments with both the restricted action space design as well as an unrestricted action space design. The restricted action space is similar to the lab experiment, and described in the previous section. In the unrestricted (or full) action space, participants can instead choose any number between 0 and 10. This enables us to evaluate whether our results hold in a setting that more closely mirrors the typical implementation of dictator and lying games.

On the second treatment dimension, we vary the *second game* used to measure spillovers by adjusting whether the DM has a personal stake or not. Specifically, we vary whether the second game is a SPECTATOR LYING-DICTATOR GAME (as before) or a regular LYING-DICTATOR GAME. It is conceivable that the spectator game has less scope for spillovers, because it is seen as a different moral context than the first party game. Including treatments with the LYING-DICTATOR GAME as the second game may therefore provide a stronger test for spillovers.

C.2.2 Remaining Procedures, Questionnaire, and Payment

We conducted the online experiment on Prolific with 1,600 participants in August 2024. We attempted to collect a sample that is representative of the general U.S. population in terms of gender, age, and ethnicity. In contrast to the lab experiment, only 5% of participants were selected to have their payments be payoff-relevant. One of the two parts was randomly selected to determine payment. The instructions included one comprehension question and one attention check. Participants who twice answered the comprehension question incorrectly or once answered the attention check incorrectly were excluded from the experiment. After completing the study, participants went through a brief questionnaire that elicited basic demographics. The median completion time was approximately 6 minutes. Participants received $1.90 for their participation on average.

C.3 Results

C.3.1 Descriptive Statistics

Of the 1,600 participants, 414, 392, 404, and 390 participants were randomly assigned to treatments RESTRICTED-FIRSTPERSON, RESTRICTED-SPECTATOR, FULLSPACE-FIRSTPERSON, and FULLSPACE-SPECTATOR, respectively. Table C.1 presents descriptive statistics. As can be seen, individual characteristics are balanced between treatments (Panel A) and random draws (Panel B).

Table C.1: Descriptive Statistics Online Experiment

Panel A. Randomization of covariates across treatments

	Overall	Restricted-FirstPerson	Restricted-Spectator	FullSpace-FirstPerson	FullSpace-Spectator	χ^2
Male	0.472	0.500	0.459	0.465	0.462	0.614
	(0.499)	(0.501)	(0.499)	(0.499)	(0.499)	
Age	45.09	44.39	46.25	44.54	45.23	0.123
	(15.80)	(15.86)	(15.23)	(16.07)	(15.99)	
Race/Ethnicity						
White	0.637	0.606	0.666	0.670	0.605	0.089
	(0.481)	(0.489)	(0.472)	(0.471)	(0.489)	
Black	0.136	0.161	0.134	0.116	0.134	0.320
	(0.343)	(0.368)	(0.341)	(0.320)	(0.341)	
Latino	0.076	0.074	0.061	0.076	0.095	0.361
	(0.265)	(0.263)	(0.239)	(0.265)	(0.293)	
Asian	0.067	0.089	0.055	0.058	0.066	0.211
	(0.251)	(0.285)	(0.229)	(0.234)	(0.248)	
Native American	0.010	0.005	0.013	0.013	0.011	0.647
	(0.101)	(0.070)	(0.114)	(0.112)	(0.102)	
College	0.603	0.603	0.618	0.610	0.579	0.707
	(0.489)	(0.490)	(0.486)	(0.488)	(0.494)	
Monthly income	4753	4601	4988	4832	4596	0.871
	(4406)	(4313)	(4436)	(4581)	(4295)	
Democrat	0.446	0.447	0.436	0.458	0.444	0.941
	(0.497)	(0.498)	(0.497)	(0.499)	(0.497)	
Observations	1600	414	392	404	390	

Panel B. Randomization of draws within treatments

	Overall	Restricted-FirstPerson	Restricted-Spectator	FullSpace-FirstPerson	FullSpace-Spectator	χ^2
HIGH	0.462	0.440	0.487	0.458	0.464	0.599
	(0.499)	(0.497)	(0.500)	(0.499)	(0.499)	
LOW	0.450	0.490	0.423	0.428	0.456	0.197
	(0.498)	(0.501)	(0.495)	(0.495)	(0.499)	
Five	0.088	0.070	0.089	0.114	0.079	0.146
	(0.284)	(0.256)	(0.286)	(0.318)	(0.271)	

Note: Panel A contains means (standard errors) of participant characteristics pooled for all participants (column 1) and the treatments (columns 2 to 5), as well as the *p*-values from Pearson's chi-squared tests (column 6). Panel B displays the shares of HIGH random draws, LOW random draws, and random draws equal to five. *College* is an indicator variable for whether a participant completed college; *Monthly* income is participants' net monthly income; *Democrat* is an indicator for whether a participant is a Democrat rather than a Republican or Independent.

C.3.2 Motive Selection

The next two sections closely follow our preregistered pre-analysis plan (see Section D.2 of the Online Appendix). As a first step, we test for moral motive selection. As in the laboratory experiment, we do so by testing whether participants are more (less) likely to tell the truth (equalize payments) after a HIGH than after a LOW random draw. We start by examining the two treatments that most closely resemble the laboratory experiment: the treatments with the restricted action space (treatments RESTRICTED-SPECTATOR and RESTRICTED-FIRSTPERSON). In these treatments, we find strong evidence of moral motive selection. In particular, 62.9% of participants choose to tell the truth after a LOW and 79.9% choose truth-telling after a HIGH random draw ($p < 0.001$). Equally, 22.0% choose equality after a LOW random draw and 11.8% choose equality after a HIGH random draw ($p < 0.001$). Note that these differences are smaller than in the laboratory experiment, perhaps because the online participants have a lower tendency to lie, reducing the scope for motive selection.

Next we turn to the treatments with the full (unrestricted) action space. These treatments allow us to test whether moral motive selection is at least partially driven by the restricted action space and the resulting high salience of moral motives. We find no evidence that this is the case. In treatments FULLSPACE-FIRSTPERSON and FULLSPACE-SPECTATOR, around 61.5% of participants tell the truth after a LOW and 83.4% choose truth-telling after a HIGH random draw ($p < 0.001$). Equally, 21.9% choose equality after a LOW draw and 12.0% choose equality after a HIGH random draw ($p < 0.001$). In other words, we still observe strong evidence of moral motive selection even when we remove the restrictions on the action space. More formally, neither the difference in truth-telling between high and low draws nor the difference in equality choices differs significantly between the restricted and full action space treatments (truth-telling: 17.0pp vs 22.1pp, $p = 0.272$; equality: 10.2pp vs 9.9pp, $p = 0.951$). In fact, only 30 of the 794 participants (around 5%) make a choice that does not directly correspond to one of the four motives. This explains the lack of a difference between the treatments with the unrestricted and restricted action space.[12]

Overall, we therefore find that participants engage in moral motive selection even if they are allowed to choose any number between 0 and 10. This provides evidence of the generality of the tendency to engage in motive selection. This finding also supports the validity of the comparisons between the LYING-DICTATOR GAME and the LYING GAME and DICTATOR GAME we focus on earlier. Moreover, the presence of moral motive selection appears to be robust to variations in the sample (student vs. general population) and type of experiment (lab vs. online).

[12] Across all random draws, truth-telling rates (equality rates) are somewhat higher (lower) in the restricted than in the full action space treatments.

C.3.3 Spillovers

To test for spillovers, we examine whether the random draw in the LYING-DICTATOR GAME affects the moral motive chosen in the second game. We first report data from the two treatments that most closely resemble the laboratory experiment, in which the second game is a spectator game (treatments RESTRICTED-SPECTATOR and FULLSPACE-SPECTATOR). Motivated reasoning predicts that participants should be less likely to tell the truth in the second game after a LOW draw in the first game (compared to a HIGH draw), and more likely to choose equality. The results are directionally in line with this hypothesis, but small and not significant. In particular, 47.1% of participants choose to tell the truth in the second stage after observing a LOW draw in stage 1. 51.6% choose truth-telling after a HIGH random draw in stage 1 ($p = 0.227$). Equally, 47.4% choose equality after a LOW random draw and 43.3% choose equality after a HIGH random draw ($p = 0.270$). Hence, we once again observe no significant spillover effects in these treatments.

When we instead examine the treatments where we replace the second game with a first-person game (treatments RESTRICTED-FIRSTPERSON and FULLSPACE-FIRSTPERSON), we find a precisely estimated null effect. In particular, 49.7% of participants tell the truth in the second stage after a LOW and 47.7% tell the truth after a HIGH random draw in stage 1 ($p = 0.576$). Equally, the share of participants choosing equality is very similar for a LOW and HIGH random draw (31.4% vs. 33.2%; $p = 0.588$). These results are directionally opposite to the motivated reasoning hypothesis, and not significant. A formal test of the difference in spillover effects between the spectator and first-person games also reveals no significant differences (truth-telling: 4.5pp vs -2.1pp; $p = 0.210$; equality: 4.1pp vs -1.8pp; $p = 0.239$).

Overall, we therefore do not find any effect of the first-stage random draw on motive choices in stage 2 (i) even if the motive choice is made before the random draw is observed and (ii) independently of whether the motive choice affects a participant's own income or not. We therefore observe no spillover effects supporting the argument that participants have stable moral preferences rather than engage in motivated reasoning.

ONLINE APPENDIX
Navigating Moral Trade-offs

Kai Barron
WZB Berlin

Robert Stüber
University of Warwick

Roel van Veldhuizen
Lund University

Contents

A Robustness Checks for Lying Game and Dictator Game **2**
A.1 Comparison to Literature 2
A.1.1 Lying Game 2
A.1.2 Dictator Game 3
A.2 Order Effects 5

B Implications of Moral Motive Selection **7**

C Experimental Instructions **9**
C.1 Laboratory Experiment 9
C.1.1 Instructions for the Lying-Dictator Game 9
C.1.2 Instructions for the Lying Game and the Dictator Game 20
C.2 Online Experiment 25

D Pre-Registrations **34**
D.1 Laboratory Experiment 34
D.1.1 Preliminaries 34
D.1.2 Main Analysis 35
D.1.3 Robustness Analysis 35
D.1.4 Additional Analysis 36
D.1.5 Power Calculations 36
D.2 Online Experiment 39
D.2.1 Main Analysis 39
D.2.2 Power Calculations 41

References **42**

A Robustness Checks for Lying Game and Dictator Game

A.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.

A.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.[1] Crucially, this implementation of the lying game allows the researcher to observe lying at the individual level. Our results are presented in Figure A.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).[2] This statistic is robust to restricting attention to the participants who observe a LOW draw and lie: of these 49 participants, only 4 (8%) lie partially and report a 5, while 45 (92%) lie maximally and

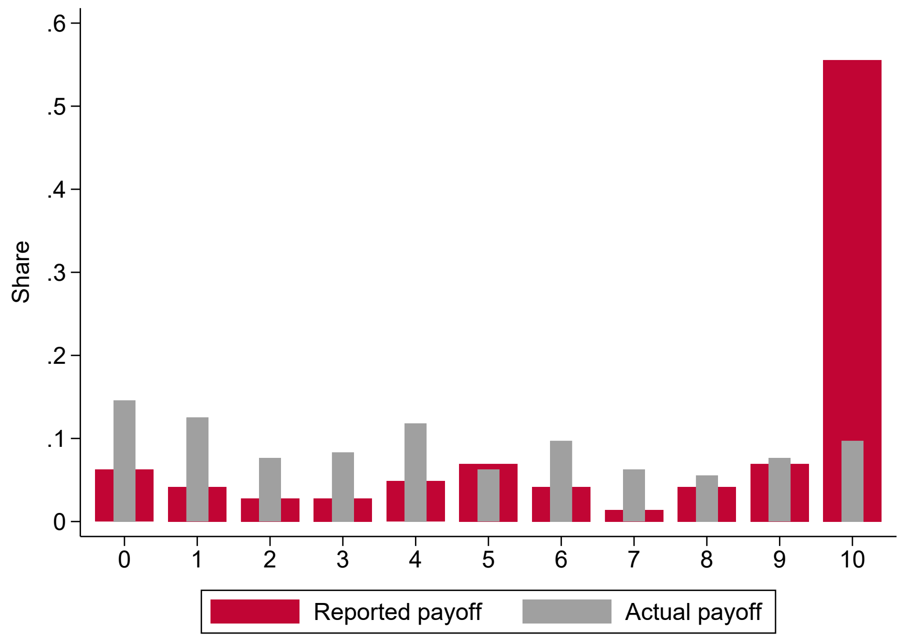
[1] One difference is that in Gneezy et al. (2018) participants can observe a number between 1 and 10, while in our study the random number varies between zero and 10 (see the *Numbers treatment* of their *Observed Game*). In our setting it is important to allow participants to be able to choose zero, 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 moral motives by restricting each participant's choice set to four items in each of the three games.

[2] 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, participants in our experiment can only lie to zero, 5, or to 10, and 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 who 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 because this is held constant across our games, we do not view it as a major concern for treatment comparisons.

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



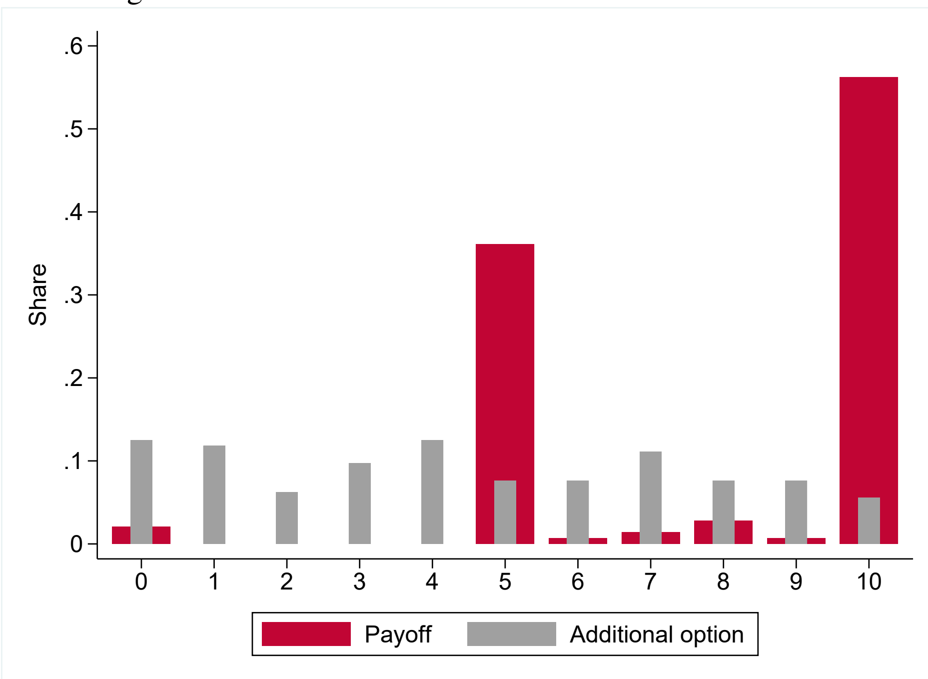
Note: The figure shows the numbers reported by participants in the Lying Game (red bars) and the actual number they observed (gray bars).

A.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 Figure 2).

Our distribution of dictator choices is shown in Figure A.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).

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



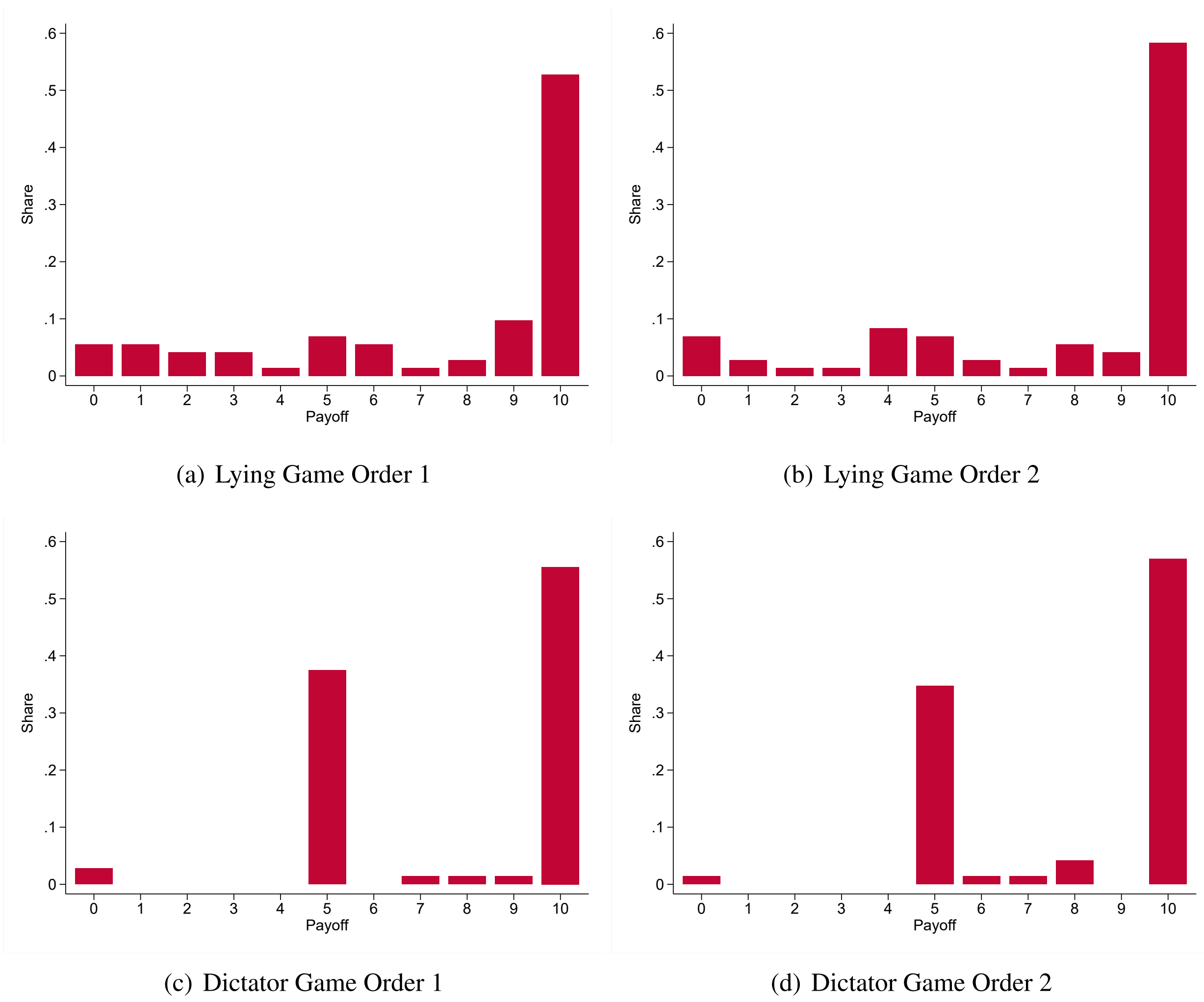
Note: The figure shows the distribution chosen by participants in the Dictator Game (red bars) and the number they could have chosen in addition to 0, 5, and 10 € (gray bars).

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 1 out of every 11 participants in our experiment can choose each of the outcomes other than 0, 5, and EUR 10. Hence, a participant who would like to keep 7, 8, or EUR 9 for themselves might end up choosing EUR 10. A similar logic can explain the increase in the proportion of participants choosing to retain 50% (36% vs. 17% in Engel, 2011). However, 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 results replicate those in the literature.

A.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.

Figure A.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).

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 A.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, $p = 0.696$).

B Implications of Moral Motive Selection

In this appendix, we describe two additional findings about how people make moral decisions when facing a trade-off between different moral motives in more detail. First, the moral 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 pool the data from the LYING-DICTATOR GAME with one of the baseline games (either LYING GAME or DICTATOR GAME) and regress a dummy for adhering to a moral motive (either truth-telling or equality) as an outcome variable on (i) an indicator for a HIGH draw, (ii) an indicator for the LYING-DICTATOR GAME, (iii) and the HIGH draw–LYING-DICTATOR GAME interaction. We therefore run four regressions: two that compare behavior in the LYING-DICTATOR GAME to behavior in the LYING GAME and two that compare behavior in the LYING-DICTATOR GAME to behavior in the DICTATOR GAME, with either a binary outcome variable that denotes choosing truth-telling or a binary variable for equality.

Table B.1 contains our results. For instance, as shown in column (1a), in the LYING GAME, participants opted to tell the truth 30pp more often after a HIGH draw than a LOW draw, as one would expect. However, the coefficient on the interaction term indicates that in the LYING-DICTATOR GAME, truth-telling was chosen 61pp more often after a HIGH draw (i.e., the responsiveness of truth-telling to the random draw is markedly higher in the LYING-DICTATOR GAME than in the LYING GAME). As can be seen in columns (1b), (2a), (2b), the interaction term of HIGH draw and LYING-DICTATOR GAME is also highly statistically significant for the other specifications.[3] Therefore, the results demonstrate that motive choices are substantially more responsive to the random draw in the LYING-DICTATOR GAME than in either of the respective baseline games. This implies 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. Whereas this comparative static of behavior can be accommodated by both the stable moral preferences and the motivated reasoning models discussed previously, it shows an increased sensitivity to the cost of adhering to a moral motive in the presence of multiple moral motives.

[3] It is worth noting that in the DICTATOR GAME the random draw does not vary the cost of adhering to any motive because the truth-telling motive is absent. In line with this, participants do not adjust their propensity to choose equality as a function of the random draw in the DICTATOR GAME (see column (2b)). However we observe a 23pp higher rate of choosing the truth-equivalent option after a HIGH draw than a LOW draw (column (2a)). The reason for this is that a subset of participants would prefer to take the selfish option over dividing the money equally, but would prefer an option that gives a small amount to the other participant over the fully selfish option. After a HIGH draw, they have this option in their choice set and so choose to report the random draw; in contrast, after a LOW draw they do not, and choose the selfish option. Importantly, however, the coefficient on the interaction term is large and statistically significant even when analyzing truth-telling in the DICTATOR GAME.

Table B.1: Elasticity of motive choices

	Comparison: Lying Game		Comparison: Dictator Game	
	Truth-Telling (1a)	Equality (1b)	Truth-Telling (2a)	Equality (2b)
LYING-DICTATOR GAME	-0.242***	0.125**	0.415***	0.123
	(0.072)	(0.048)	(0.071)	(0.086)
HIGH	0.299***	0.232***	-0.051*	0.044
	(0.084)	(0.059)	(0.025)	(0.085)
LYING-DICTATOR GAME x HIGH	0.313**	0.380***	-0.326***	-0.420***
	(0.112)	(0.095)	(0.080)	(0.114)
Constant	0.380***	0.013	0.051*	0.342***
	(0.055)	(0.013)	(0.025)	(0.055)
Observations	247	247	249	249

Note: OLS estimates, robust standard errors in parentheses. Dependent variable is a dummy variable for choosing to tell the truth (columns a) or equalizing payments (columns b). "LYING-DICTATOR GAME" is a dummy variable for the LYING-DICTATOR GAME, "HIGH" is a dummy variable for randomly drawing a high (> 5) number, and the final variable is the interaction of these two variables. The comparison group is the baseline game the behavior in the LYING-DICTATOR GAME is compared to. Subjects who fulfill multiple motives in the lying or dictator game are coded as telling the truth. *** p<0.01, ** p<0.05, * p<0.1.

Second, selfish behavior is less common in the LYING-DICTATOR GAME than in the two baseline games. When pooled across LOW and HIGH draws, the share 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 participants choose the selfish option ($p < 0.002$ and $p < 0.001$ for the comparisons with the two games). Yet, the decision makers' payoffs do not differ significantly across the three games. They are 7.26 in the LYING-DICTATOR GAME, 7.58 in the LYING GAME, and 7.85 in the DICTATOR GAME ($p = 0.388$ and $p = 0.064$ for the comparisons with the DICTATOR GAME and the LYING GAME, respectively). This is because in the LYING-DICTATOR GAME the reduced propensity to choose the selfish option (which reduces payoffs) and the ability to engage in moral motive selection (which increases payoffs) offset one another on average. The implication is that in the LYING-DICTATOR GAME the average individual is able to achieve the same payoffs but has a higher propensity of complying with at least one moral principle. Essentially, being moral becomes cheaper when additional moral motives are introduced, provided one is willing to choose which moral motive to adhere to.

C Experimental Instructions

C.1 Laboratory Experiment

This appendix contains the experimental instructions for treatment LYING-DICTATOR (Appendix C.1.1) and treatment BASELINE, that is, the LYING GAME and DICTATOR GAME (Appendix C.1.2). Parts in *italic* were not shown to the participants. Original instructions were in German.

C.1.1 Instructions for the Lying-Dictator Game

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

D.1.1.1 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:[4]

- 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

[4]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.[5] 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.

[5]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[6], before you receive your payment.

Questionnaire

Screen 16:[7] **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)?

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

[7]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.

D.1.1.2 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).

C.1.2 Instructions for the Lying Game and the 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:[8]

- 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.

[8]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 €	5 €
○ 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.

C.2 Online Experiment

This appendix contains the experimental instructions for the online experiment. The instructions for treatments *Restricted-FirstPerson* and *Restricted-Spectator*, and of treatments *FullSpace-FirstPerson* and *FullSpace-Spectator* are equivalent in Part 1. The instructions for treatments *Restricted-FirstPerson* and *FullSpace-FirstPerson*, and of treatments *Restricted-Spectator* and *FullSpace-Spectator* are equivalent in Part 2. Parts in *italic* were not shown to the participants.

Screen 1: Welcome

You are participating in a study on decision making. The study takes up to **10 minutes** to complete. You will receive **$1.90 (£1.50)** for completing the study. In addition, you may be able to earn a **bonus payment**.

The study is anonymous. Hence, your identity will not be revealed to others, and the identity of others will not be revealed to you.

Next, you will see the instructions. **Please read the instructions carefully as they describe how you can earn the bonus payment**. You will be asked questions to confirm that you have read the instructions. If you answer these questions incorrectly, you will be excluded from the study and won't be eligible for payment.

This study is approved by the Ethics Committee of WZB Berlin under the condition that there is no deception. **Hence, all the information that we provide is true**.

By continuing to the next screen, you consent to participate in this study. For more details about your consent, click on "See consent form."

Consent

You have been invited to take part in a study to learn more about how individuals make decisions in common scenarios.

Dr. Kai Barron (WZB Berlin), Dr. Robert Stüber (NYU Abu Dhabi), and Prof. van Veldhuizen (Lund University) are conducting this study. The purpose of this study is to further our understanding of how people make decisions.

Participation in this study is voluntary, and you must be at least 18 years of age to participate. You may choose to withdraw at any time. There are no known risks associated with your participation in this project beyond those of everyday life. Other than the compensation amount you agreed upon before you entered into this survey, there are no direct benefits from participation.

Your answers are anonymous and will not be shared with anyone. The confidentiality of these research records and your identity will be strictly maintained. To preserve the anonymity of the data, the researchers do not know your identity, and their data is never associated with any personal

identifiers. All data generated will be stored on the researcher's computer.
If there is anything unclear or if you have questions, you may contact Dr. Robert Stueber at robert.stueber@nyu.edu.

Screen 2: General instructions

This study consists of **2 parts** and a questionnaire.
At the end of the study, a computer algorithm will randomly select **5%** of participants. For each selected participant, **one** of the two parts will be randomly selected. If you are one of the selected participants, all of your choices in that part will be implemented. This means that they will affect your bonus payments and the bonus payments of other individuals as described.

Understanding Question

If you are one of the selected participants, which statement below is true?

- There are two parts. **Both parts** will be used to calculate your bonus payment.
- There are two parts. **One part** will be randomly selected to calculate your bonus payment.
- There are two parts. **Part 1** will be used to calculate your bonus payment.

Screen 2 Repeated: General instructions

Your answer to the understanding question was incorrect. Please re-read the instructions and answer the understanding question once again. This is your last opportunity to get the right answer.
This study consists of **2 parts** and a questionnaire.
At the end of the study, a computer algorithm will randomly select **5%** of participants. For each selected participant, **one** of the two parts will be randomly selected. If you are one of the selected participants, all of your choices in that part will be implemented. This means that they will affect your bonus payments and the bonus payments of other individuals as described.

Understanding Question

Please answer the question below. If you answer incorrectly, you will be excluded from the study.
If you are one of the selected participants, which statement below is true?

- There are two parts. **Both parts** will be used to calculate your bonus payment.
- There are two parts. **One part** will be randomly selected to calculate your bonus payment.
- There are two parts. **Part 1** will be used to calculate your bonus payment.

Part 1: Lying-Dictator Game
Treatments Restricted-FirstPerson and Restricted-Spectator

Screen 3: Instructions Part One

In part one of the study 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. That is, the actual order may be 9, 2, 4, ...; 3, 6, 1; ... or any other 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 will determine how $10 is divided between you and another Prolific participant. Therefore, it will determine your bonus payment **and** the bonus payment of the other participant.

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:[9]

- 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."

You can make any of the four reports you prefer. Your report will not be linked to your identity or influence your submission status on Prolific in any way.
After you have made your decision, you will be asked to briefly write an explanation for why you chose your selected report.

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

If you are one of the selected participants and this part is randomly selected, your decision will determine how $10 is divided between you **and** the other participant.

Treatments FullSpace-FirstPerson and FullSpace-Spectator

Screen 3: Instructions Part One

In part one of the study 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. That is, the actual order may be 9, 2, 4, ...; 3, 6, 1; ... or any other 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 will determine how $10 is divided between you and another Prolific participant. Therefore, it will determine your bonus payment **and** the bonus payment of the other participant.

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 have to make your report. You can do so in several ways:[10]

- 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."
- Choose any other number and report: "The number I saw was [number reported]."

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

You can make any of the eleven reports you prefer. Your report will not be linked to your identity or influence your submission status on Prolific in any way.
After you have made your decision, you will be asked to briefly write an explanation for why you chose your selected report.

If you are one of the selected participants and this part is randomly selected, your decision will determine how $10 is divided between you **and** the other participant.

Screen 4:

Please click on one of the boxes:

Screen 5:

Treatments Restricted-FirstPerson and Restricted-Spectator

Please choose the report you would like to make.

Out of the four reports, you can freely choose the report that you prefer.

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

Treatments FullSpace-FirstPerson and FullSpace-Spectator

Please choose the report you would like to make.

Out of the eleven reports, you can freely choose the report that you prefer.[11]

○ Report "The number I saw was **0**." You receive **$0**. Other participant receives **$10**. (Maximize the other participant's payment)
○ Report "The number I saw was **1**." You receive **$1**. Other participant receives **$9**.
○ Report "The number I saw was **2**." You receive **$2**. Other participant receives **$8**.
○ Report "The number I saw was **3**." You receive **$3**. Other participant receives **$7**.
○ Report "The number I saw was **4**." You receive **$4**. Other participant receives **$6**.
○ Report "The number I saw was **5**." You receive **$5**. Other participant receives **$5**. (Equalize payments)
○ Report "The number I saw was **6**." You receive **$6**. Other participant receives **$4**.
○ Report "The number I saw was **7**." You receive **$7**. Other participant receives **$3**
○ Report "The number I saw was **8**." You receive **$8**. Other participant receives **$2**.
○ Report "The number I saw was **9**." You receive **$9**. Other participant receives **$1**. (Tell the truth)
○ Report "The number I saw was **10**." You receive **$10**. Other participant receives **$0**. (Maximize your payment)

Screen 6:

Treatments Restricted-FirstPerson and Restricted-Spectator

[11] Here, we show an example for when the randomly drawn number was 9.

You chose the following option: Tell the truth and report: “The number I saw was d.”
Please briefly explain why you have chosen this option.

Treatments FullSpace-FirstPerson and FullSpace-Spectator
You chose the following report: “The number I saw was ...”
Please briefly explain why you have chosen this report.

Part 2: Spectator Lying-Dictator Game

Screen 7:

Treatments Restricted-FirstPerson and FullSpace-FirstPerson

Instructions Part Two

In part two of the study, you will play the same game as before with one key difference: you must decide how to make your report **before** seeing which random number you draw.

That is, before you click on a box to see the random number, you must choose one of the following ways to make your report:

- Tell the truth and report: “The number I saw was [number you will see].”
- 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 choice will apply to any number you see. After making your choice, you will again see 11 boxes containing the numbers 0 to 10 hidden inside. You will then click on a box to see the number. The computer will then implement your decision about how to make your report.

If you are one of the selected participants and this part is randomly selected, your decision will determine how $10 is divided between you and another participant.

Treatments Restricted-Spectator and FullSpace-Spectator

Instructions Part Two

In part two of the study, you will play a similar game as before with two key differences:

1. The number you report to us **does not determine** your bonus payment. Instead, it determines the bonus payment of **two other participants** in this study, that we label Participant A and Participant B. Note that these two participants **did not** make a decision in the first part of the study and **will not** make any decision in this study. In other words, these participants’ earnings depend on your decision.

2. You must decide how to make your report **<u>before</u>** seeing which random number you draw. That is, before you click on a box to see the random number, you must choose one of the following ways to make your report:

   - Tell the truth and report: “The number I saw was [number you will see].”
   - 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.”

   Your choice will apply to any number you see. After making your choice, you will again see 11 boxes containing the numbers 0 to 10 hidden inside. You will then click on a box to see the number. The computer will then implement your decision about how to make your report.

If you are one of the selected participants and this part is randomly selected, your decision will determine how $10 is divided between Participant A and Participant B.

Screen 8:

Treatments Restricted-FirstPerson and FullSpace-FirstPerson

Please choose the report you would like to make.

Your choice will apply to any number you reveal.

○ Tell the truth and report: “The number I saw was [number you will see].” You receive $[number]. Other participant receives $[10-number].
○ Equalize payments and report: “The number I saw was **5**.” You receive **$5**. Other participant receives **$5**.
○ Maximize your payment and report: “The number I saw was **10**.” You receive **$10**. Other participant receives **$0**.
○ Maximize other participant’s payment and report: “The number I saw was **0**.” You receive **$0**. Other participant receives **$10**.

Treatments Restricted-Spectator and FullSpace-Spectator

Please choose the report you would like to make.

Your choice will apply to any number you reveal.

○ Tell the truth and report: “The number I saw was [number you will see].” Participant A receives $[number]. Participant B receives $[10-number].
○ Equalize payments and report: “The number I saw was 5.” Participant A receives $5. Participant B receives $5.
○ Maximize Participant A’s payment and report: “The number I saw was 10.” Participant A receives $10. Participant B receives $0.
○ Maximize Participant B’s payment and report: “The number I saw was 0.” Participant A receives $0. Participant B receives $10.

Screen 9:

Please click on one of the boxes:

Screen 10:

Treatments Restricted-FirstPerson and FullSpace-FirstPerson

Given your choice and the number you draw, you report: "The number I saw was ..." You receive ... Other participant receives ...

Treatments Restricted-Spectator and FullSpace-Spectator

Given your choice and the number you draw, you report: "The number I saw was ..." Participant A receives ... Participant B receives ...

Screen 11:

You are done with **Parts 1 and 2**, please answer a few questions about yourself.

Screen 12:

Data quality is important. We would like to know whether you are still paying attention rather than just clicking "Continue." Please now choose the third option from the top (forty-one to sixty percent) and ignore the question below. What is the share? (0-20%, 21-40%, 41-60%, 61-80%, 81-100%)

Screen 13: Demographic questions

You are almost done, please answer the following questions.

What is your age?

Which gender do you identify with? (male, female, other)

What is your race/ethnicity? Select all that apply.

In which state do you currently reside?

What is your highest level of education?

Screen 14: Final questions

To finish, please answer the final set of questions.

Generally speaking, do you think of yourself as a: (Democrat, Independent, Republican, Other)

What is your employment status? (Full-time employed (30 hours a week or more); Part-time employed (less than 30 hours a week); Taking care of home or family; Student; Retiree; Unemployed; Prefer not to say)

What was your personal income before taxes over the past year? Please include income from all sources. If you are not sure, please enter your best guess. (Less than $1,000$ per month, $12,000$ per year; $1,000-2,000$ per month, $12,000-24,000$ per year; $2,000-3,000$ per month, $24,000-36,000$ per year; $2,000-3,000$ per month, $24,000-36,000$ per year; $3,000-4,000$ per month, $36,000-48,000$ per year; $4,000-5,000$ per month, $48,000-60,000$ per year;

$5,000-7,000$ per month, $60,000-84,000$ per year; $7,000-10,000$ per month, $84,000-100,000$ per year; $10,000-15,000$ per month, $120,000-180,000$ per year; $15,000-25,000$ per month, $180,000-300,000$ per year; More than $25,000$ per month, $300,000$ per year; Prefer not to say)

Screen 15:

Was there anything that was unclear during the study?

Screen 16: Thank you for participating!

You have reached the end of this study.

You will receive **$1.90 ($1.50)** for completing the study.

If you earn a positive bonus payment, you will receive it in a couple of days once all study participants have finished.

Please do not discuss the procedures or content of this study with anyone. Note that you can only participate in this study once. If we find multiple participants with the same Prolific ID then we count only the first participation.

If you have any further questions, please contact us at robert.stueber@nyu.edu.

Prolific ID

Please confirm your Prolific ID. (**Important**: make sure it is correct or we may not be able to pay you!)

It is IMPORTANT that you click on the button below to record your answers and ensure you receive your payment.

D Pre-Registrations

In this section, we reprint the pre-analysis plans and power calculations that we included in our pre-registrations.[12] For the pre-analysis plans, we also note where the results of the pre-registered specifications can be found in the paper. Note that for the laboratory experient we only pre-registered the LYING-DICTATOR treatment. The primary purpose of the BASELINE treatment was to serve as a comparison for the patterns observed in the LYING-DICTATOR GAME.

D.1 Laboratory Experiment

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.

[12]The lab experiment was preregistered at the AEA registryas 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. The online experiment was preregistered at the AEA registry as trial no. AEARCTR-0013827 and can be found here: https://doi.org/10.1257/rct.13827-1.0

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 2.2.1. 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 2.2.2.

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 2.2.2 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 2.2.2.

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 2.2.2.

D.1.5 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. Because 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.

D.2 Online Experiment

We conducted the online experiment after circulating an earlier version of our paper featuring just the laboratory experiment. This allowed us to base our pre-analysis plan directly on our previous working paper.

D.2.1 Main Analysis

We will start by replicating our main analysis from the earlier version of the experiment, as reported in our existing *working paper*:

1. Motive Selection: does the random draw affect the game 1 report?

   (a) We will report the fraction of participants in the first game who choose (i) truth-telling, (ii) equality, (iii) payoff maximization, or (iv) any other report as a function of the random draw in the first game (either HIGH or LOW as per D1a above).

   (b) We will test whether the truth-telling (equality) rate in the first game is significantly higher (lower) for HIGH rather than LOW random draws. In both cases, we will use a two-sided test of proportions.

   (c) We will conduct these analyses (E1a and E1b) separately for each of the two action space treatments (full vs restricted) while pooling across the other treatment dimension (i.e., pooling the treatments with SLDG and LDG in game 2).

Authors' notes: These tests are presented in Appendix C.2.2.

2. Spillovers: does the game 1 random draw affect the game 2 report?

   (a) We will report the fraction of participants in the second game who choose (i) truth-telling, (ii) equality, (iii) payoff maximization (in the LDG only), or (iv) any other report as a function of the random draw in the first game (either HIGH or LOW as per D1a above).

   (b) We will test whether truth-telling is chosen more often than equality in the second game for HIGH random draws from the first game. We will test whether equality is chosen more often than truth-telling in the second game for LOW random draws from the first game. We will use a two-sided test of proportions in both cases.

(c) We will do this analysis (E1a and E1b) for each of the two spillover treatments (LDG vs SLDG), pooling across the other treatment dimension (i.e., pooling the two action spaces).

Authors' notes: These tests are presented in Appendix C.2.3.

In addition, we will also look at the following treatment differences:

3. Action Space: does restricting the action space affect motive selection (E1)?

   (a) We will test whether the frequency of each type of report (truth-telling, equality, payoff maximization, or other) differs between the full action space and restricted action space treatments, separately for LOW and HIGH random draws (per D1a). We will use a two-sided test of proportions in each case, and pool across the two treatments within each treatment arm (i.e., pool the treatments with SLDG and LDG in game 2).

   (b) We will regress truth-telling (1-truth-telling, 0-other reports) on a dummy for HIGH draws (HIGH-1, LOW-0), a dummy for the full action space treatments (1-full, 0-restricted), and the interaction of the two. We will then test whether the interaction term is significant using a two-sided test. We will pool within the action space treatment dimension (i.e., for each action space treatment, we will pool the treatments with SLDG and LDG in game 2).

   (c) We will repeat the previous test with equality (1-equality, 0-other reports) as the dependent variable instead.

   (d) In case the frequency of "other" reports (i) differs by more than 15 percentage points and (ii) differs significantly by treatment (1-full, 0-restricted), we will also redo analysis E3b and E3c while removing participants that chose the "other" report.

Authors' notes: These tests are presented in Section C.2.2. The frequency of "other" reports does not differ by more than 15 percentage points between treatments.

4. Spillovers and Game Type: does the spillover effect (E2) depend on the game type (SLDG vs LDG)?

   (a) We will regress truth-telling (1-truth-telling, 0-other reports) in the second game on a dummy for HIGH draws (HIGH-1, LOW-0) from the first game, a dummy for the game type (1-SLDG, 0-LDG) and the interaction of the two. We then test whether the interaction term is significant using a two-sided test.

(b) We will repeat the previous test with equality (1-equality, 0-other reports) as the dependent variable instead.

(c) If there is no significant effect of action space in E3, we will pool the data from both action space arms. If there is a significant effect of action space (at the 5% level in any of tests E3a-3c) we will instead conduct tests E4a and E4b separately in each treatment arm.

Authors' notes: These tests are presented in Section C.2.3. We provide the results pooling the data as well as separately for each treatment arm.

D.2.2 Power Calculations

We will aim to collect 400 participants per treatment, which would give us a power of 0.80 to detect motive selection or spillover effects of around 10pp.

1. Motive Selection (E1):

In our original experiment, we observed a 61pp increase in truth-telling (from 14% to 75%) and a 38pp decrease in equality choices (from 47% to 9%) for HIGH draws (relative to LOW draws). We expect smaller effect sizes in our online sample. Within each treatment arm (restricted vs full), our intended sample size (800 participants) would give us approximately 363 participants with LOW and HIGH draws, respectively (taking into account that we do not use participants with a draw of 5). This would allow us to detect effect sizes of approximately 12pp with a power of 0.9 and 10pp with a power of 0.8, respectively (e.g., "power twoproportions 0.4, n(726) power(0.9)" in Stata). Smaller effect sizes are detectable for proportions that lie closer to 0 or 1.

2. Spillovers (E2):

For the test that pools the data within each action space treatment arm, our power would be identical to the previous test.

3. Action Space (E3):

Here, we rely on a Difference-in-Difference test examining whether the difference in e.g., truth-telling rates between HIGH and LOW draws varies depending on the action space. Assuming an effect size of 15pp (HIGH 0.575 vs LOW 0.425) in the restricted action space treatments, we would have a power of approximately 0.80 to detect whether this effect size vanishes in the full action space treatments (with a sample size of 800 in each treatment arm). Hence, with the proposed sample size, we could reliably detect whether a motive selection effect of around 15pp vanishes completely after introducing a full action space.

4. Game Type (E4):

For the test pooling the data within each game type arm (both types of action spaces), the power would be the same as in part E3 above. If we look separately at each type of action space, we would need an effect size of around 20pp to achieve a similar kind of power.

References

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

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

FEHR, E. AND K. M. SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, 114, 817–868.

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

Discussion Papers of the Research Area Markets and Choice 2024

Research Unit: **Market Behavior**

Christian Basteck An axiomatization of the random priority rule	SP II 2024-201
Leonardo Felli The emergence of enforcement	SP II 2024-202
Steffen Huck, Hans-Theo Normann, Fidel Petros Concerns about rising prices may raise prices	SP II 2024-203

Research Unit: **Economics of Change**

Kai Barron, Robert Stüber, Roel van Veldhuizen Navigating moral trade-offs	SP II 2024-301

All discussion papers are downloadable:
http://www.wzb.eu/en/publications/discussion-papers/markets-and-choice