University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 291

The Big Robber Game

Carlos Alós-Ferrer, Jaume García-Segarra and Alexander Ritschel

June 2018

The Big Robber Game*

Carlos Alós-Ferrer† University of Zurich

Jaume García-Segarra‡ University of Cologne

Alexander Ritschel§ University of Zurich

This Version: June 2018

Abstract

We present a novel design measuring a correlate of social preferences in a high-stakes setting. In the Big Robber Game, a "robber" can obtain large personal gains by appropriating the gains of a large group of "victims" as seen in recent corporate scandals. We observed that more than half of all robbers take as much as possible. At the same time, participants displayed standard, prosocial behavior in the Dictator, Ultimatum, and Trust games. That is, prosocial behavior in the small is compatible with highly selfish actions in the large, and the essence of corporate scandals can be reproduced in the laboratory even with a standard student sample. We show that this apparent contradiction is actually consistent with received social-preference models. In agreement with this view, in the experiment more selfish robbers also behaved more selfishly in other games and in a donation question. We conclude that social preferences are compatible with rampant selfishness in high-impact decisions affecting a large group.

JEL Classification: C72 · C92 · D03

Keywords: Big Robber Game · Social preferences · Corporate scandals · Incentives

*We thank Larry Blume, Chetan Dave, Ernst Fehr, Urs Fischbacher, Pablo Hernández-Lagos, Georg Kirchsteiger, Eliran Halali, Nikos Nikiforakis, Ernesto Reuben, Klaus Ritzberger, Stefan Schulz-Hardt, and Matthias Sutter for helpful comments and discussions. Alexander Ritschel was financed by the Research Unit "Psychoeconomics" (FOR 1882; Project AL-1169/4) of the German Research Foundation (DFG). Icons used in Figure 1 were created by Freepik from `www.flaticon.com`.

†Corresponding author: carlos.alos-ferrer@econ.uzh.ch. Department of Economics, University of Zurich. Blümlisalpstrasse 10, 8006 Zurich, Switzerland.

‡Department of Economics, University of Cologne. Albertus-Magnus-Platz, 50923 Cologne, Germany.

§Department of Economics, University of Zurich. Blümlisalpstrasse 10, 8006 Zurich, Switzerland.

1 Introduction

The recent decades have witnessed an astonishing loss of confidence by the general public on financial institutions, large firms, and economic decision makers. The image of riotous protesters at WTO ministerial conferences, G20 summits, and other high-level meetings has become commonplace. Public animosity toward financial institutions, the banking sector, and "Wall Street" reached a 40-year high in 2011 (Owens, 2012). Beyond the circles of academic economists and politicians, confidence in the market has been replaced by the views put forward by former German Chancellor Helmut Schmidt, who coined the term "predatory capitalism" (*Raubtierkapitalismus*; Schmidt, 2003), referring for instance to a large number of scandals where corporate executive officers were viewed as unduly appropriating funds or actively damaging society. As stated by Krugman (2008), "[Americans have] lost confidence in the integrity of our economic institutions." This has sparked a widespread, cynical view of economic actors as hopelessly-selfish agents capable of inflicting any damage on others for personal gain.[1]

On the other hand, it is clear that self-interest is not the only motivation guiding economic human decisions, and this has long been recognized in economics. As Adam Smith put it, "how selfish soever man may be supposed to be, there are evidently some principles in his nature, which interest him in the fate of others, and render their happiness necessary to him, though he derive nothing from it, except the pleasure of seeing it" (Smith, 1759). The limits of self-interest have been systematically exposed in experiments using stylized games as the Ultimatum Game (Güth et al., 1982), the Dictator Game (Forsythe et al., 1994), and the Trust Game (Berg et al., 1995), as well as in many distributive-allocation experiments (Engelmann and Strobel, 2004). This evidence has motivated models incorporating various forms of "social preferences," encompassing fairness concerns, prosociality, and even motivations with an intentional component as reciprocity (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000; Charness and Rabin, 2002; Dufwenberg and Kirchsteiger, 2004; Falk and Fischbacher, 2006; Alger and Weibull, 2013).

The "social-preference revolution" seems to be based on the enticingly simple message that economic agents are not so selfish as usually assumed in economic theory. For example, in a widely-echoed article, Mazar et al. (2008) argue that most people will help others, refrain from actively damaging them, and, if given opportunity, cheat only a little. This message, however, is at odds with the popular (and populist) criticisms of the "economic system," which showcase the allegedly inordinate levels of selfishness displayed by corporate executives and decision makers in the financial sector. In other words, while social-preference models attempt to defuse the assumption that economic agents are selfish, the popular perception of economic decision makers is that, at least at certain levels, selfishness is rampant.

[1]The idea that corporate scandals might have even macroeconomic consequences is reflected, for instance, in the model of Myerson (2012), where the possibility of moral-hazard rents extracted by bankers and financial agents can create macroeconomic credit cycles with repeated recessions.

In view of these opposed trends, it is maybe advisable to make a distinction between social preferences and the experimental observation of prosocial or altruistic behavior. To drive this point home, consider the following example. A decision maker $i = 1$ can choose among certain allocations for n participants, including herself, $x = (x_1, \ldots, x_n)$, with the property that $\sum_{i=1}^{n} x_i = C$ for a certain amount $C > 0$. Her utility function is given by

$$u(x) = x_1 - 25\left(\frac{x_1}{C} - \frac{1}{n}\right)^2.$$

This is an example of social preferences as in Bolton and Ockenfels (2000). Suppose this decision maker participates in a Dictator Game where she can freely allocate 10 monetary units between herself and a second participant, that is, $n = 2$, $C = 10$, $0 \leq x_1 \leq 10$. A straightforward computation shows that u is maximized at $x_1 = 7$, hence the decision maker will freely donate 3 out of 10 monetary units to the other player, in line with prosocial behavior as typically observed for the Dictator Game in the lab. That is, this decision maker can be seen as prosocial. Suppose that this very same decision maker is now part of a group of 17 people, each of them endowed with 10 monetary units, and she is given the power to appropriate any amount τ from each of the other participants (uniformly). That is, $n = 17$, $C = 170$, and the decision maker can freely choose among allocations of the form $\big(10 + 16\tau, 10 - \tau, \overset{16}{\ldots}, 10 - \tau\big)$, and hence her problem consists of maximizing

$$v(\tau) = 10 + 16\tau - 25\left(\frac{10 + 16\tau}{170} - \frac{1}{17}\right)^2$$

over $0 \leq \tau \leq 10$. A direct computation shows that $v(\cdot)$ is strictly increasing in its domain, and hence our prosocial decision maker will decide $\tau = 10$, appropriating all income from a group of 16 other participants.

In our view, it is uncontested that, first, humans care about how their actions affect others and, second, selfishness is one of the most fundamental and powerful human motivations. The difference between prosocial behavior as observed in the lab and rampant egoism as apparently observed in the field is that the situations triggering the respective responses are fundamentally different. As the example above illustrates, this is readily captured by extant models of social preferences. In this work, we aim to demonstrate empirically that prosocial behavior in the small is fully compatible with morally-outrageous behavior in high-stakes, high-impact decisions where a decision maker can obtain a large personal gain at the expense of significantly damaging a large number of other people. Obtaining such empirical evidence entails three kinds of difficulties, and it is our aim to address them within an experimental paradigm designed to inform the discussion.

First, many of the experimental paradigms on which social-preference models have been based are constrained to bilateral interactions. This is natural, since those games were a first step in research, and keeping the experimental setting as simple as possible facilitates the analysis. Think, however, of the typical situation that critics of financial institutions have in mind, where a single decision maker has the possibility to harm a

large number of people. Clearly, such a situation differs from the experimental paradigms in several dimensions, and hence the latter are not well-suited to predict behavior in the former. For example, in view of corporate scandals, it could be argued that financial institutions select people who are not representative of the general population, and hence prosocial behavior in the large is compatible with selfish behavior at certain levels. Currently-employed experimental paradigms, however, are not appropriate to test for such a selection phenomenon, since one would need an experimental setting where regular laboratory subjects can actually damage a group of other experimental participants (in exchange for significant monetary gain). This is exactly what our design provides.

The second, related difficulty is that games of the bargaining type (Ultimatum, Dictator, and Trust games) are highly stylized and might hence measure a very specific dimension of behavior. Again, this is natural, since they were originally designed to provide demonstrations of the existence of social preferences and motives. But they might be insufficient to explore the full impact of prosocial behavior on economic decisions. In other words, those games might be triggering prosocial behavior in a limited set of situations, which might be compatible with high levels of selfishness being sparked in different circumstances (see also Levitt and List, 2007). Our experiment contributes to the literature supporting this view.

The third difficulty is that, to determine whether prosocial behavior is of sufficient magnitude as to be seen as a major driving force for economic behavior, one needs to address the size of the stakes. This question is hard to examine in standard laboratory experiments, simply because the stakes are too low. There are, however, indications that prosocial behavior can still be found in spite of high stakes. Indeed, experiments in low-income countries have found prosocial behavior with large stakes (Roth et al., 1991; Slonim and Roth, 1998; Andersen et al., 2011).[2] However, an analogous, smoking-gun demonstration for high-income, developed countries is missing, because generating high stakes in such environments would quickly exhaust even the most generous research budget. Our design makes a methodological contribution by providing a cost-neutral way (in terms of research budget expenditure) to study high-stake settings.

In this contribution, we present a new experimental paradigm, the "Big Robber Game," which adds to the debate on prosociality vs. selfishness along the three dimensions sketched above. First, it focuses on a situation where a single decision maker can take a significant amount of money *from a large group of other participants*, and is hence closer to phenomena sparking public-opinion concerns in recent years. Second, it provides a qualitatively different paradigm which, together with standard games as the Ultimatum, Dictator, and Trust games, allows for a more complete exploration of the nature and consequences of social preferences. Third, it allows us to test for prosocial behavior by educated subjects in developed countries, with reasonably high stakes (around

[2]Andersen et al. (2011) focused on responder behavior in Ultimatum games and showed that rejection rates decreased with high stakes, but were still positive.

100 Dollars/Euros), in a natural, straightforward frame where being selfish inflict serious damage on a group of others.

The basic, pragmatic design idea behind the Big Robber Game is as follows. Consider an average researcher in economics based in the U.S. or Europe, with access to a budget for behavioral experiments. A single session with, say, 30 participants, costs around €/$ 400 (or more). From the point of view of the average participant, the cost of that individual session is a significant amount of money. Suppose that the experimenter selects a participant randomly from half of the participants in the session and gives him or her the possibility of taking as much as 50% of the earnings of the other half of the players in the session. Then, a player in this game will face a single decision potentially giving him or her the possibility to walk away with €/$ 100. That decision might seriously damage a relatively large number of people and be considered selfish, even antisocial. The experimental session, however, does not cost more. Our experimenter will spend exactly the same part of his or her research budget as a "regular" experiment would have required.

In our design, we allocated participants to two possible roles, which were framed neutrally (type I and II) but which we will refer to here as "robbers" and "victims." Robbers were asked the Big Robber question, that is, which fraction of the earnings of the victims they wished to take away. Crucially, only one robber was selected at the end (and his or her decision actually implemented), but all robbers had to answer the question in advance (as in the strategy method; Selten, 1967; Brandts and Charness, 2011). All victims were informed in advance of the possibility that a fraction of their earnings could be taken away by a robber. There was only one Big Robber decision, with fixed roles, that is, robbers could not be robbed themselves, and there was no possible retaliation. To generate the relevant income, players of both roles provided their decisions for the active and passive roles in a series of intermediate games: Dictator, Ultimatum, and Trust games. This allows us to compare behavior in the Big Robber question to behavior in these standard games, and hence link the different measures of prosociality that they provide. Feedback was not provided until the end of the experiment, and participants did not know the identity or any characteristic of the players they would be matched with in the intermediate games (not even if they were robbers or victims). Therefore, behavior could not be affected by possible group rivalry (see, e.g., Abbink et al., 2010, 2012).

Beyond the primary research question, we were also interested to see whether merely asking the Big Robber question might have an effect on subsequent behavior, since a number of psychological theories (see Section 4 below) might predict behavioral effects. Hence, we further divided the set of robbers in two, creating two different treatments. "Ex ante robbers" were asked the Big Robber question at the beginning of the experiment, before they provided their decisions for the intermediate games. "Ex post robbers" were asked the Big Robber question *after* they had made their decisions for the inter-

mediate games (note that, since only victims could be robbed, the Big Robber question did not affect the robbers' earnings from the intermediate games).

After all decisions had been made, but before they were implemented and payoffs were determined, we asked participants whether they would donate a percentage of their earnings (net of the show-up fee) to a local charity. They were free to specify any fraction. The idea of this "donation question" was to provide an independent measure of possible guilt, by giving robbers the chance to give a prosocial use to their appropriated earnings.

The theoretical predictions of the selfish *homo oeconomicus* for the Big Robber Game are straightforward: take as much as possible. Received social-preference models predict varying levels of selfish behavior, which we examine in Section 6. Those models incorporate parameters of prosociality which can differ across subjects. The intermediate games provide exactly such measures of prosociality. We use the Dictator Game to calibrate the models of Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002), and Alger and Weibull (2013). As in our example above, the predictions for these models suggest that the agent should in many cases take as much as possible, i.e. 50%; actually, standard social-preference models predict higher levels of selfish behavior than we actually observe.

The results of the experiment show that the paradigm provides new insights into social preferences and prosocial behavior which go beyond the ones that can be gathered with "standard" games as the Dictator, Ultimatum, and Trust games. First, at the stakes level of the Big Robber Game, there is a considerable degree of "selfishness" as predicted, with as much as 56% of robbers deciding to take as much as allowed of the victims' earnings. However, the behavior of the very same robbers in the Dictator, Ultimatum, and Trust games is within standard ranges. This shows that the Big Robber Game provides indeed a novel view for the debate on prosociality vs. selfishness, and that different situations even within the same experiment can trigger qualitatively different behavior with respect to the perceived prosociality associated with the decisions.

Second, we do find that robbers who decide to take as much as possible behave more selfishly than robbers who do not take the maximum in the Dictator, Ultimatum, and Trust games, and also in the donation question. That is, the Big Robber Game is not orthogonal to other paradigms; on the contrary, behavior in our paradigm does correlate with individual prosociality, and hence is consistent with underlying social preferences as modeled in the literature. Third, we do find some (weak) treatment effects between ex ante and ex post robbers which are compatible with known psychological theories. However, these effects interact with gender, showing basic differences in the effect that revealing oneself to be selfish has on subsequent behavior. Gender differences were not unexpected, and hence we took care to have a perfectly balanced sample, with exactly half of all robbers within each treatment being female.

Our study is of course related to several branches of the extensive literature on social preferences. Some experimental paradigms in this literature allow players to take money from other players. In the Power-to-Take game (Bosman and van Winden, 2002;

Bosman et al., 2005; Reuben and van Winden, 2010), a player can attempt to take the earnings of another player, but the second player can react by destroying his own income and hence the part appropriated by the first player. Players in the "take role" choose considerable take rates in this paradigm. Bosman et al. (2006) consider a version with group decisions and show that behavior is very similar to the individual version. In the Moonlightning Game (Abbink et al., 2000), the first mover can either transfer money to a second mover or steal from him, but the second mover is able to punish. Results showed consistent punishment of stealing behavior although punishment was costly. Bardsley (2008) considered a Dictator Game with the added option to take income away from the recipient (see also List, 2007), and showed that Dictator giving is greatly reduced when the option to take exists. Andreoni (1995) found more selfish behavior when taking from a public account than when giving to a public good. Khadjavi and Lange (2015) replicated the experiment making both the taking and giving options simultaneously available and found that allowing for taking reduces giving. These results are qualitatively in agreement with ours and show that situations where taking is possible generally trigger less prosocial behavior than those framed in terms of bargaining or sharing.

The literature has also examined a number of paradigms where antisocial behavior emerges in the laboratory, for instance where participants can destroy the earnings of others or of a group of innocents, due to envy or a concern for relative payoffs (Zizzo and Oswald, 2001; Zizzo, 2004; Abbink and Sadrieh, 2009; Abbink and Herrmann, 2011; Karakostas and Zizzo, 2016). Among those, we single out the experiment of Zizzo (2004), where participants could appropriate other participants' income or transfer it to others. Decisions were made within small (four-player) groups where everybody could be robbed by the other three participants. This creates a situation closer to a social dilemma where a participants' act of stealing could be argued to be self-defense, since other participants will most likely steal from him or her.

Our paradigm goes beyond previous insights by providing a (relatively) high-stakes situation where an individual can take up to half of the earnings of a large group of other participants, while systematically comparing decisions to simultaneous Dictator, Ultimatum, and Trust games. In contrast with the paradigms mentioned above, our focus is on decisions affecting a large group (not bilateral interactions), with high stakes involved, and where the decision to take the earnings of others is non-strategic, that is, can be made with impunity.

Conceptually, our study is aligned with Levitt and List (2007), who argue that behavior is crucially linked to not only the underlying preferences but also the specific characteristics and properties of the situation at hand. The Big Robber Game provides a specific, relevant setting where qualitative behavior becomes predominantly selfish, even though the same participants simultaneously display standard levels of prosocial behavior in Dictator, Ultimatum, and Trust games played during the very same experimental session.

The paper is structured as follows. Section 2 presents the experimental design in detail. Section 3 discusses the results of the Big Robber Game in itself, that is, the decision to take away (or not) the earnings of other players. It also examines the differences in behavior (in the intermediate games) across robbers who behave more or less selfishly. Section 4 discusses treatment effects, that is, differences in behavior due to whether the Big Robber question had already been answered or not when players faced the intermediate games. Section 5 examines the answers to the final donation question. Section 6 examines the predictions of well-known models of social preferences for the Big Robber Game, performing a simple out-of-sample estimation analysis. Section 7 concludes. The Appendix presents the experimental instructions.

2 Experimental Design

2.1 Design and Procedures

The objective of the design was to allow for the measurement of a correlate of social preferences where significant personal gain can be obtained at the cost of economically harming a large group of fellow participants. The novel design of the experiment allowed us to accomplish this objective while maintaining an affordable total experiment cost, even though relatively high stakes were involved and the experiment was run in a developed, high-income country (Germany).

The experiment consisted of three parts: the Big Robber question, a sequence of games (Dictator, Ultimatum, and Trust games), which we will refer to as the *intermediate games*, and a final block of questionnaires, including a belief elicitation question and the opportunity to donate a fraction of the individual earnings to a charity organization. There were two between-subject treatments, which differed in the timing of the Big Robber question. In the *ex ante treatment*, the Big Robber question was asked before participants played the sequence of games. In the *ex post treatment*, the order was the reverse one. In the latter treatment, robbers were not aware of the possibility of robbing until after they provided all their decisions for the sequence of games. This design serves two purposes. First, it allows us to disentangle robbing behavior from possible confounds due to the order of tasks and whether the involved income has already been generated or not. Second, by comparing across treatments, it becomes feasible to test for possible effects of the Big Robber question on subsequent behavior.

We now discuss the three parts in detail. Figure 1 presents an overview of the design. At the very beginning, participants were informed that the experiment consisted of multiple parts but no details were provided.

The Big Robber question revealed to the participants that there were two types of players in the experiment. There were 32 participants per session. In each session, 16 participants were randomly assigned to the robber role and the remaining 16 to the victim role. The instructions referred to these roles as "Type I" and "Type II," but we will call

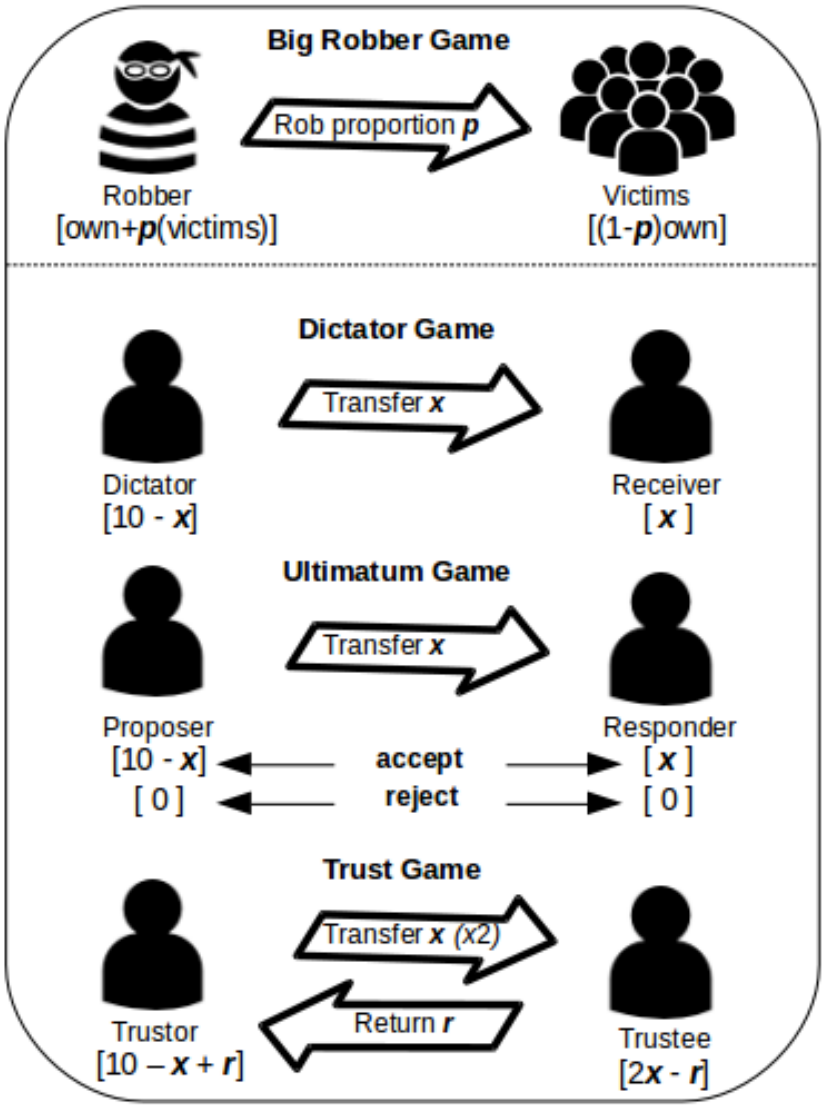

Figure 1: Overview of the Experimental Design.

them robbers and victims for concreteness. The robbers were informed that, at the end of the experiment, one of them would be randomly selected and would have the chance to take 50%, 33%, 10%, or nothing from the joint earnings of all victims. It was clear that only victims would be robbed, that is, a robber would never be robbed. Every robber was then asked to provide his or her decision, should he or she be selected (as in the strategy method). In this way, we collected answers to the Big Robber question from all robbers. This question was asked either before or after the sequence of games depending on the treatment. Specifically, in the ex post treatment the robbers were initially not provided with any information about the existence of the Big Robber question. In contrast, in both treatments, before the sequence of games started, the victims were informed of the fact that participants of the other type would be asked the Big Robber question, and that one of them would be randomly selected and his or her decision implemented.

Table 1: Big Robber Choices.

Choice	Proportion in points	Amount in Euros
☐	50% of all points	about €100 (expected, maximum €112)
☐	33% of all points	about €66 (expected, maximum €74.60)
☐	10% of all points	about €20 (expected, maximum €22.40)
☐	0% of all points	€0

To emphasize the stakes, the Big Robber question also included numerical estimates of the amount of money that would be transferred to the randomly selected robber depending on each possible answer. Those numerical estimates were computed using estimates of behavior in the sequence of games taken from the literature (details are provided below), which delivered an estimated transfer of €100 for a robber deciding to take 50% (respectively €66, €20, and €0 for 33%, 10%, and 0%). We also provided the theoretical maximum transfers derived from the joint profit-maximizing choices. The exact format of the Big Robber choices and the provided information is displayed in Table 1.

The sequence of intermediate games was played after the Big Robber question in the ex ante treatment and before it in the ex post treatment. The order of those games was fixed and identical for all participants. Each participant played first the Dictator Game (Forsythe et al., 1994) as a dictator, an Ultimatum Game (Güth et al., 1982) as proposer, and a Trust Game (Berg et al., 1995) as trustor. Next, he or she "played" the Dictator Game as receiver, that is, the participant was simply informed that at the end of the game he or she would be matched to some other participant and play the passive receiver role in a Dictator Game. Then, the participant played the Ultimatum Game as responder and, finally, the Trust Game as trustee. Decisions for the last two roles were collected using the strategy method, that is, participants provided contingent answers for each possible proposal of the sender (proposer or trustor, respectively). The order of the games was chosen to minimize the changes between sender and receiver mindsets, hence avoiding an artificial activation of the concept of reciprocity through mere alternation. The framing was in terms of "decisions" and the roles were described neutrally (participants A and B for sender and receiver roles, respectively). Payoffs were realized at the very end of the experiment by randomly matching participants in the corresponding roles in such a way that every participant was matched to a different participant for each of the six games. In particular, no feedback was provided for any of the decisions until the experiment ended. There was no distinction between robbers and victims for the purposes of matching to implement the six intermediate games.

Each game was played with an endowment of 10 points. That is, in each of the games, the sender (dictator, proposer, or trustor) decided how to split 10 points among him- or herself and the other player (with only integer allocations allowed). In the Dictator Game(s), this decision was implemented. In the Ultimatum Game(s), the responder

decided whether to accept or reject each possible split (proposal). At the end of the experiment, after matching players and implementing decisions, an acceptance led to the realization of the proposal, and a rejection led to zero payoffs (in that particular game) for both proposer and responder. In the Trust Game, the trustor proposal was implemented, but the part allocated to the trustee was then doubled and the trustee could decide which part of the proceeds (if at all, and constrained to be an integer number of points) to send back to the trustor. The factor of two in the Trust Game(s) was chosen to reduce artificial incentives to cooperate and capture the social preferences in a clean way (Glaeser et al., 2000).

The purpose of the sequence of games was twofold. On the one hand, as most of the literature we use them as (standard) indicators of the social preferences of the participants. On the other hand, the games were a device to generate income for the victims which could then be partially taken away by the robbers.

As commented above, we estimated the income generated by the sequence of games in order to provide numerical estimates of the additional revenue associated to each alternative choice in the Big Robber question. Since each game was played with an initial endowment of 10 points and the amount sent by the trustor in the Trust Game was doubled, the maximum number of points that could be earned for each two players was 80, resulting in a maximum of $32 \times 40 = 1280$ points for the whole session, of which half (640) corresponds to the victims. With an exchange rate of 35 Eurocents per point, this means that by robbing 50% the selected robber would receive a maximum transfer of €112, and proportionally for the other alternatives, as reported in Table 1. To estimate the *expected* transfers, we relied on the literature. We computed the expected earnings relying on the meta-analyses of Oosterbeek et al. (2004) for the Ultimatum Game and Johnson and Mislin (2011) for the Trust Game.[3] In this way, we obtained proxies for the expected behavior in the sequence of games and were able to compute a numerical estimate of €195.17 for the joint earnings of the victims, leading to estimates (rounding up) of approximately €100, €66, and €20 for the decision to rob 50%, 33%, and 10%, respectively (as reflected in Table 1). The actual average joint earnings of victims in our sessions was €179.34, showing that our estimate was reasonably accurate.

The third and last part of the experiment consisted in a block of individual questions. First, participants were asked about their beliefs regarding the Big Robber question. Specifically, they were requested to indicate how many out of 100 people they believed

[3]The joint earnings in a 10-point Dictator Game are obviously constant and equal to 10. According to Oosterbeek et al. (2004), the rejection rate in implementations of the Ultimatum Game played in Germany is about 9.5%, hence the expected joint earnings in a 10-point Ultimatum Game are 9.05 points(10 points with a 90.5% acceptance rate). Johnson and Mislin (2011) contains a Trust Game experiment carried out in Germany with a factor of two, and where the trustor sent on average 58% of the initial endowment (Walkowitz et al., 2009). Hence the expected joint gains (which are independent of the decision of the trustee) for an initial endowment of 10 points are $4.2 + 2 \times 5.8 = 15.8$. This leads to a total of $2 \times (10 + 9.05 + 15.8) = 69.7$ for each set of six two-player games, or $(1/2) \times 16 \times 69.7 = 557.6$ for all 16 victims in a session. With an exchange rate of 35 Eurocents per point, this translates into €195.17, hence stealing 50% would result in an average transfer of €97.59, which we rounded up to €100.

would take 50%, 33%, 10%, or 0%. This belief elicitation question was not incentivized. Second, the participants were given the opportunity to donate a fraction of their total earnings (excluding the show-up fee) to a local charity organization. This was done to examine possible guilt feelings on the part of the robbers. We will discuss this question in Section 5 below. Last, the experiment included a questionnaire on demographic data (including gender, age, and field of studies).

After all three parts of the experiment were completed, payoffs were computed. First, the matching algorithm paired the decisions of participants in the six games and computed the generated income. Then, the "Big Robber" was randomly chosen among the 16 robbers and his or her decision was implemented. Then, the donation decision was individually applied to each participant's earnings. Finally, all payoff-relevant information was presented to the participants. Payment was made anonymously (individually) in a separate room.

2.2 Power Analysis

Before running the experiment, we conducted a power analysis to determine the sample size. We focused on possible effects in the Dictator Game, and specifically the hypothesis that facing the robbing decision before the Dictator Game might affect the dictator's decision. As a prior, we used the joint distribution of the meta-analysis of behavior in the Dictator Game by Engel (2011). Starting from this distribution we used three different treatment effect sizes and simulated how many observations would be needed to find a significant difference between the prior distribution and the distribution influenced by the robbing decision. We simulated distributions with deviations of 15, 25, and 30 percentage points from the prior at the focal points represented by transfers of 0 and 5 points (out of 10). Between these values we smoothed the distribution.

Of course, the distribution might be different across groups of participants making different robbing decisions. For example, participants taking 50% are likely to be more selfish than those taking 0%. We expected that robbing decisions would concentrate in two groups, with the remaining categories capturing only a small share.[4] In addition, we expected gender effects, and hence invited enough participants of each gender to ensure an equal amount of male and female robbers. Accounting for these factors, we computed that, with 40 observations for each gender and group of decisions in each treatment, a medium-sized treatment effect ($\approx 25\%$) would still have a power of 80% when restricting to a specific gender. Assuming evenly-split decisions, this means 160 observations per treatment, with 80 men and 80 women each, for a total group size of 320 robbers (and hence 320 victims).

[4] An alternative would have been to conduct a pretest to determine base rates for the robbing decision. However, given the nature of this experiment, we decided not to conduct any pretest in order to prevent the possibility that future participants at our lab might have heard about the design.

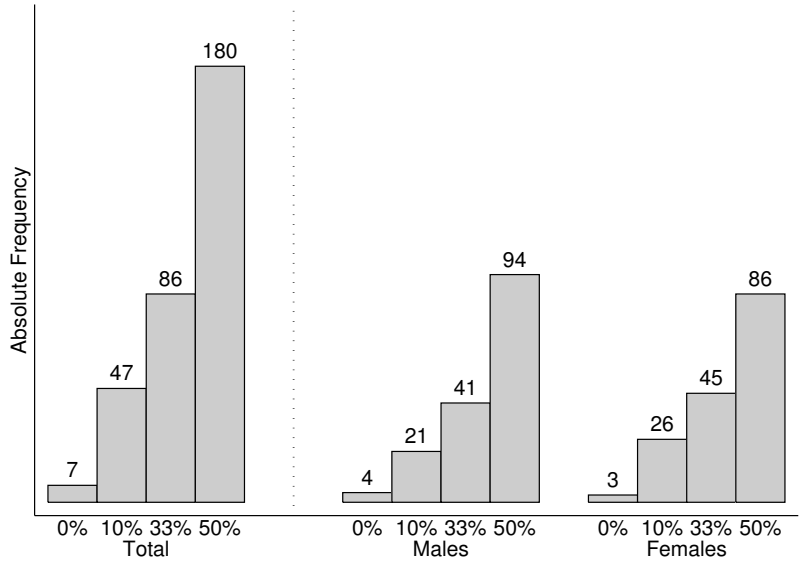

Figure 2: Absolute Frequency of Robbing Choices. The left-hand side shows the robbing decisions pooled for males and females. The right-hand side shows the robbing decisions split by gender.

2.3 Data

We conducted the experiment at the Cologne Laboratory for Economic Research (CLER). The experiment was programmed in z-Tree (Fischbacher, 2007) and participants were recruited from the student population of the University of Cologne using ORSEE (Greiner, 2015), excluding psychology students. There were 640 participants (334 females). We collected data in 20 sessions scheduled in four consecutive days, 10 according to the ex ante treatment, and 10 according to the ex post treatment.

Subjects belonging to the group of "robbers" were perfectly balanced by gender for each treatment and each session, resulting in a total of 80 male and 80 female ex ante robbers, and 80 male and 80 female ex post robbers. There were 72 male and 88 female victims in the ex ante sessions and 74 male and 86 female victims in the ex post sessions.

The average payoff from the six intermediate games was €11.15 (around $12 at the time of the experiment), ranging from €4.20 to €19.60. Additionally, participants received a show-up fee of €2.50. Experimental sessions lasted around 50 minutes.

3 The Big Robber Question

3.1 Robbers' Choices: To Rob or Not to Rob

The left-hand side of Figure 2 depicts all robbing decisions. A majority of robbers opted for personal gain at the expense of the 16 victims in their session. Out of 320 robbers, 180 (56.25% of the robbers) decided to steal the maximum, i.e. 50% of the victims' earnings, while 86 further robbers (26.88%) decided to take 33%, only 47 robbers (14.69%) took

just 10% (the minimum above zero), and a purely anecdotal 2.19% (just 7 robbers) declined to take anything.

The right-hand side of Figure 2 depicts the robbing decisions conditional on gender. There is almost no difference in the robber choices between males and females. Out of 160 male robbers, 94 (58.75%) took 50%, 41 (25.62%) took 33%, 21 (13.12%) took just 10%, and only 4 (2.50%) declined to take money from the victims. Out of 160 female robbers, 86 (53.75%) took 50%, 45 (28.12%) took 33%, 26 (16.25%) took just 10%, and only 3 (1.88%) declined to rob. The distributions are not significantly different according to a χ^2 test ($\chi^2(3) = 1.216$, $p = 0.7491$).[5]

After the Big Robber was selected, there was actual stealing in all 20 sessions. The average robbing earnings (i.e. transfer due to the robbing decision) was €66.83, ranging from €17.85 to €97.65. Counting the earnings from the games, the 20 selected Big Robbers earned an average of €78.23, ranging from €26.25 to €110.25 (not including the show-up fee and the donation decisions). Hence, the experiment truthfully involved high stakes, since it was actually possible to walk away with over €100.[6]

Since the vast majority of robbers took either 50% or 33%, for the subsequent analysis we divided the robbers in two groups. The "more selfish" ones are those who decided to rob 50%, hereafter denoted as *robbers-50%*. The "less selfish ones" are those who took less than 50%, that is, all those who decided to rob 33%, 10%, or nothing at all. We refer to this category as the *robbers-no-50%*.

The left-hand-side of Figure 3 depicts the relative frequency of robbers-50% in the ex ante and ex post treatments (recall that there are 160 observations in each treatment). In the ex ante treatment 62.50% of the robbers (100) took 50%, but in the ex post treatment only 50.00% of the robbers (80) took 50%. The difference is significant according to a test of proportions ($z = 2.254$, $p = 0.0242$).[7] The right-hand side of the figure depicts the robbing decisions conditional on treatment. The distributions are significantly different according to a χ^2 test ($\chi^2(3) = 9.157$, $p = 0.0273$). We can conclude that if the Big Robber question is asked after the games have been played, there is a small shift away from robbing the maximum. However, as we expected the treatment difference is not large and it does not detract from the observation that stealing the maximum is the dominant mode of behavior.

Nevertheless, there are natural avenues of explanation for the treatment difference. A psychologically-motivated explanation might point out that ex post robbers could have developed some empathy towards the victims after playing the 6 games. A more

[5]Of the 320 robbers in our experiment, 147 reported having economics-related majors. There were no significant differences between the distributions of robbing choices for them and for the 173 robbers who reported non-economics-related majors.

[6]Taking into account the effects of the robbing decision and the show-up fee, robbers earned on average €17.78 (median €13.70, $SD = 17.73$, ranging from €6.70 to €112.75), and victims €9.69 (median €9.50, $SD = 2.20$, ranging from €5.65 to €16.85).

[7]The difference misses significance if looking only at males: 63.75% (51) of male robbers took 50% in the ex ante treatment, versus 53.75% (43) in the ex post treatment (test of proportions, $z = 1.285$, $p = 0.1989$). Looking at females, the difference is weakly significant: 61.25% (49) took 50% in the ex ante treatment, versus 46.25% (37) in the ex post treatment (test of proportions, $z = 1.903$, $p = 0.0571$).

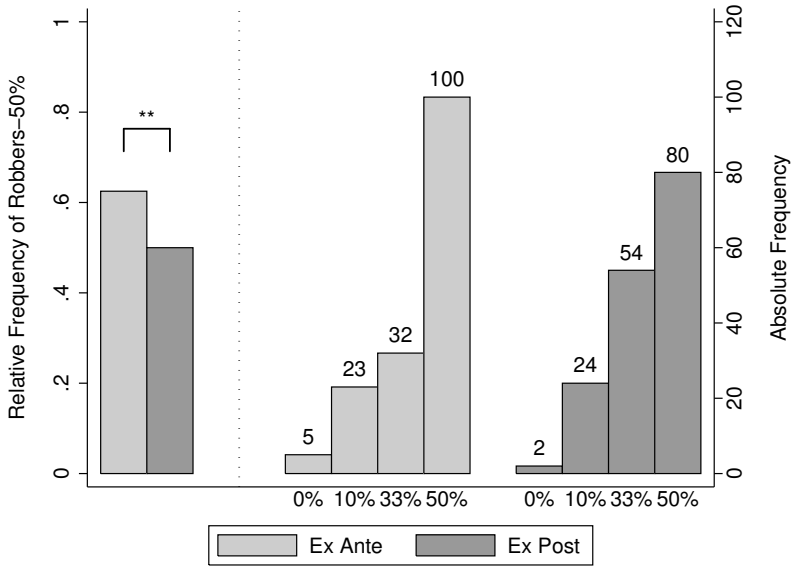

Figure 3: Robbing Choices by Treatment. The left-hand side shows relative frequency of robbers-50% by treatment. The right-hand side shows the complete histograms of robbing decisions split by treatment. ** $p < 0.05$, test of proportions.

economically-based explanation might point at income effects and attitudes toward risk. First, in comparison to ex ante robbers, ex post robbers were explicitly aware of at least part of the earnings from the games (e.g., they knew that the dictator and the trustor decisions would be implemented as given) and could have an expectation on the total earnings from the games. An increased awareness of those earnings is in practice an income effect which could decrease the motivation to steal. Second, since ex ante robbers had not experienced the six games when making the robbing decision, they could have a less-focused expectation on the earnings accruing from them, i.e. face a more risky prospect regarding those than the ex post robbers. Under risk aversion, this could lead to a larger appropriation decision, in practice compensating for the higher variance of earnings. A related explanation might be that ex ante robbers who took the maximum might be trying to insure themselves against the possibility of having low earnings in the intermediate games, while ex post robbers already had a personal estimate of possible earnings. However, in this case one could argue that robbers who took the maximum in the ex ante treatment should have made more selfish decisions in the Dictator Game than their ex post counterparts, since in this game they have actual control of their earnings, and this hypothesis is not confirmed by the data (see Section 4.4.1 below).

3.2 Decision Times

It is a long-standing, well-established observation that decisions where the decision-maker is "closer to indifference" (informally speaking) are harder and result in longer decision times (Dashiell, 1937; Mosteller and Nogee, 1951; Moyer and Landauer, 1967).

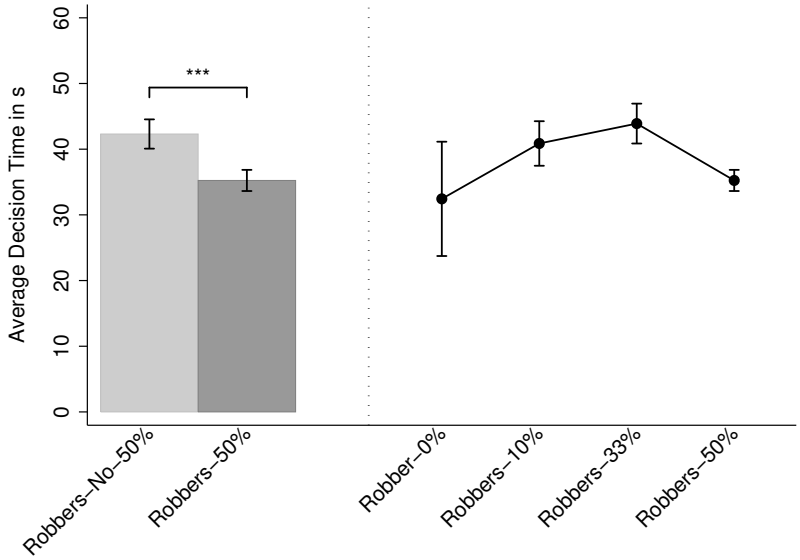

Figure 4: Decision Times for the Big Robber Question. The left-hand side shows the average decision time of robbers-no-50% versus robbers-50%. The right-hand side shows the average decision time of the robbing decision by robbing choice. The bars represent one standard error of the mean. *** $p < 0.01$, MWW test.

A recent application to economic data has been provided by Krajbich et al. (2015). The basic observation resonates with the intuition that when alternatives are similarly desirable, the decision maker will be more likely to struggle with the decision and require more time to select a choice.

We relied on this logic to investigate the possible moral struggle faced by robbers. To this end, we recorded the decision times for the Big Robber question. Initially, robbers received an explanation on the robbing decision, detailing the possibility to take part of the victims' earnings and explaining that a decision on how much to take would have to be made. As they clicked a continuation button, the table with the four possible alternatives was revealed (recall Table 1). On this table, robbers had to select a choice and then click a confirmation button. Decision times measure the time elapsed from the appearance of the table with the four alternatives to the clicking of the confirmation button. Hence they include the time needed for reading and understanding the table and making a decision.

Robbers who took the maximum decided significantly faster than other robbers. The mean decision time of robbers-50% ($N = 180$) was 35.24 s, against 42.31 s for robbers-no-50% ($N = 140$). Decision times are clearly significantly different according to a Mann-Whitney-Wilcoxon (MWW) test ($z = -2.613$, $p = 0.0090$).[8] Average decision times are depicted in the left-hand side of Figure 4.

[8]The result remains true if we compare robbers-50% only to those who took 33% ($N = 86$, mean decision time 43.90 s; MWW test, $z = -2.568$, $p = 0.0102$).

The immediate interpretation is that the more selfish robbers faced a less severe moral struggle than the ones who decided not to rob the maximum. The right-hand side of Figure 4 depicts the decision times across the different robbing decisions. We observe an inverted U-shape which agrees with this interpretation. More extreme decisions (robbing the maximum or robbing only a little) should reflect a more clear preference, hence shorter decision times, while intermediate decisions might be the result of compromising and balancing tradeoffs (the decision maker is "closer to indifference"), resulting in longer decision times. However, the extreme data point corresponding to purely altruistic behavior (declining to rob) is of course purely anecdotal since there were almost no such observations.

The difference in decision times is also clearly significant when looking at the ex post treatment alone (robbers-50%, $N = 80$, average 35.64 s; robbers-no-50%, $N = 80$, average 45.53 s; MWW test, $z = -2.503$, $p = 0.0123$). However, although the direction persists, the difference is not significant for the ex ante treatment (robbers-50%, $N = 100$, average 34.92 s; robbers-no-50%, $N = 60$, average 38.02 s; MWW test, $z = -0.652$, $p = 0.5144$).

Decision times can also be used to study whether the processes underlying a certain decision are more or less intuitive or deliberative in nature, following dual-process theories (see, e.g., Alós-Ferrer and Strack, 2014, for a recent overview) which indicate that intuitive processes are faster.[9] Following this logic, Cappelen et al. (2016) observed that fair decisions in a Dictator Game took on average 38.4 s, whereas selfish decisions took 48.5 s. Relying on this statement, i.e. that fair decisions are faster, they argue that fair decisions might be more intuitive. If we followed the argument of Cappelen et al. (2016), we would have to conclude that the decision to rob as much as possible is more intuitive than the decision to refrain from such behavior. However, decisions with decision times as long as the ones studied here or in Cappelen et al. (2016) clearly always include a large amount of deliberation, and are hence not well-suited for the study of underlying processes, at least in a straightforward way. Inferences of process characteristics in these cases risk a reverse-inference fallacy, i.e. "intuitive" may mean "fast," but this would not imply that "faster" means "more intuitive" (Myrseth and Wollbrant, 2016). Hence, we favor the simpler interpretation that faster decisions for the robbers who took the maximum indicate a reduced moral struggle in comparison with the robbers who partially overrode their impulse to take the maximum.

3.3 Beliefs

We elicited the participants' beliefs by asking how many out of 100 participants they thought would choose to take 0%, 10%, 33%, and 50% of the victims' earnings. From the resulting distributions, we computed the average percentage that each individual thought would be taken away from the victims. This average is increasing with the

[9] See Achtziger and Alós-Ferrer (2014), Alós-Ferrer et al. (2016), and Alós-Ferrer and Ritschel (2018) for examples of response-times studies of individual decisions in economic settings.

type of robber. The few robbers-0% (N=7) believed that on average only 16.73% of the victims' earnings would be taken away. The averages increased to 26.41% for the robbers-10% (N=47), 33.21% for the robbers-33% (N=86), and 41.33% for the robbers-50% (N=180). There is a clear positive correlation between the robbing choice and the beliefs ($\rho = 0.5928$, $p < 0.0001$). The result persists for each gender (males, $\rho = 0.6233$, $p < 0.0001$; females, $\rho = 0.5658$, $p < 0.0001$) and treatment (ex ante, $\rho = 0.5198$, $p < 0.0001$; ex post, $\rho = 0.6804$, $p < 0.0001$). This finding suggests that participants had beliefs consistent with their actual behavior, that is, participants believed that, on average, other participants' behavior would be close to their own.

3.4 The Selfishness of Big Robbers

It is important to establish that the Big Robber Game measures social preferences and not a completely different construct. To this end, the intermediate games corresponded to the standard ones usually employed to study social preferences, i.e., the Dictator, Ultimatum, and Trust games.

Dictator Game. The cleanest measurement of social preferences as studied in the literature is provided by the Dictator Game, since there are no strategic concerns for the dictator decision. For this reason, we placed this decision at the very beginning of the block containing the intermediate games, ensuring that there would be no income effects or other carryover considerations at this point. Robbers who took the maximum were significantly less generous as dictators in the Dictator Game. Robbers-no-50% (N=140) sent on average 3.157 points while robbers-50% (N=180) only sent 1.267 points (see the left-hand extreme of Figure 5). The difference is highly significant according to a Mann-Whitney-Wilcoxon test ($z = 8.210$, $p < 0.0001$). The result persists when splitting the data by gender or treatment. Figure 6 displays the full distribution of dictators' decisions, separately for robbers-no-50% and robbers-50%. The two distributions are significantly different ($\chi^2(8) = 79.853$, $p < 0.0001$). We conclude that robbers who took the maximum are more selfish, as measured by the Dictator Game, than other robbers.

Ultimatum Game. Participants played the Ultimatum Game twice, once as proposer and once as responder. Of course, offers in the Ultimatum Game were larger than in the Dictator Game, due to the strategic aspect of the proposer's decision (avoiding rejection). However, again robbers who took the maximum were revealed to be more selfish than other robbers, this time by their behavior as proposers in the Ultimatum Game. Robbers-50% ($N = 180$) offered 3.922 points on average while robbers-no-50% ($N = 140$) offered 4.550 points (second decision illustrated in Figure 5). The difference is highly significant (MWW test, $z = 4.141$, $p < 0.0001$). The effect persists when splitting the data by gender or treatment.

We elicited responder behavior in the Ultimatum Game using the strategy method, yielding 11 decisions per participant (whether to accept or reject proposals of 0,1,...,10

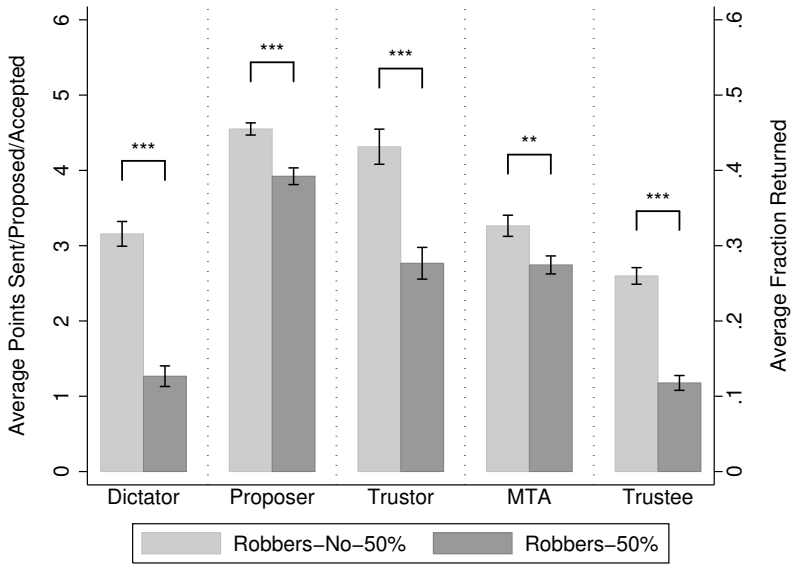

Figure 5: Robbers' Decisions in the Intermediate Games. Comparison of decisions of robbers-50% and robbers-no-50% in the Dictator Game, the Ultimatum Game, and the Trust Game. *** $p < 0.01$, ** $p < 0.05$, MWW test

points out of 10). To analyze responder behavior, we computed the smallest accepted offer (or Minimum Threshold of Acceptance, MTA), excluding participants who switched between rejection and acceptance multiple times.[10] Robbers-50% ($N = 178$) had an average MTA of 2.730 points while robbers-no-50% ($N = 135$) had a significantly higher MTA of 3.163 points (MWW test, $z = 2.437$, $p = 0.0148$; fourth decision illustrated in Figure 5). As ought to be expected, robbers who took the maximum are willing to accept lower offers since a higher level of selfishness goes hand-in-hand with giving priority to purely monetary concerns. The behavior of responders in the Ultimatum Game again indicates that for robbers who took the maximum, social preferences are less marked than for other robbers.

Trust Game. Participants played the Trust Game twice, once as trustor and once as trustee. The decision made as trustor was already the third down the line in the block of intermediate games. However, there were again clear differences, with robbers who took the maximum trusting less than other robbers. Robbers-no-50% ($N = 140$) sent on average 4.314 points to the trustee while robbers-50% ($N = 180$) sent on average only 2.767 points. This difference is highly significant (MWW test, $z = 5.636$, $p < 0.0001$; third decision illustrated in Figure 5). The effect persists when conditioning on either treatment or gender. Once again, this is evidence that social preferences are less marked for robbers who took the maximum, in the sense that they trust less, presumably because they expect *others* to be more selfish.

[10] Out of 640 participants only 19 participants (2.97%) switched more than once.

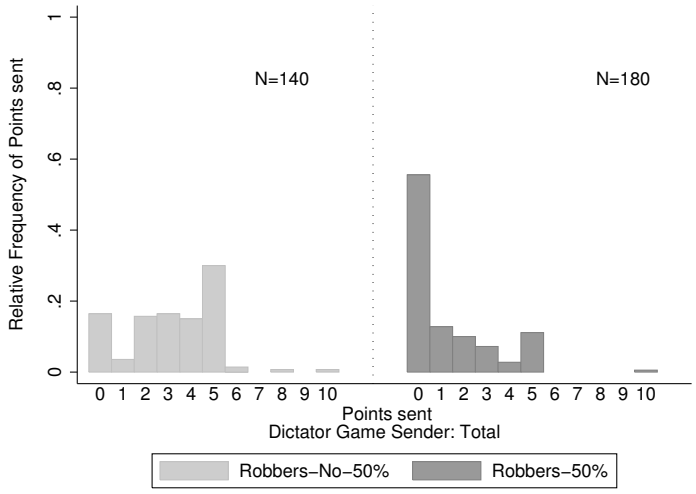

Figure 6: Behavior of Robbers in the Dictator Game. Relative frequency of points sent by the robbers-no-50% and robbers-50% in the Dictator Game

We elicited trustee behavior in the Trust Game using again the strategy method, yielding 10 decisions per participant (how much to send back if the trustor sent 1,. . . ,10 points out of 10). To analyze trustee behavior, we computed the fraction that was sent back aggregated over all 10 decisions. We again find clear differences. Robbers-no-50% (N=140) returned on average 25.99% of the received points, against only 11.77% for robbers-50%. This difference is highly significant (MWW test, $z = 8.995$, $p < 0.0001$; right-most decision illustrated in Figure 5). The effect also persists when conditioning on gender or treatment. Strictly speaking, the decision of the trustee is free of strategic components and is formally equivalent to a Dictator Game. Hence, again we see that robbers who took the maximum behave more selfishly than other robbers.

Average behavior of robbers. At the same time, behavior in the Dictator, Ultimatum, and Trust games was well within the standard ranges reported in the literature. The average offer by robbers in the Dictator Game was 20.94%, which is quite close to the grand average of 24.7% reported by Engel (2011) for students. The average offer by robbers in the Ultimatum Game was 41.97%, which was rather close to the grand average of 40.54% (71 studies) reported by Oosterbeek et al. (2004). The average MTA of robbers in the Ultimatum Game was 29.17%, which is within the ranges reported in other Ultimatum Game experiments using the strategy method. For example, Cappelletti et al. (2011) report mean MTAs of 23.00% and 31.42% for an endowment of €15 and €7, respectively. Declerck et al. (2009) found MTAs of 23.10% and 30.00% when subjects played an Ultimatum Game after and before being matched with another subject, respectively.[11]

[11] Armantier (2006) finds slightly higher mean MTAs, ranging from 31.80% to 38.00%. McLeish and Oxoby (2011) find slightly lower MTAs of 26.10% in the baseline condition of their experiment.

For the Trust Game, it is harder to compare our data to the literature because published experiments have a higher design variance than those employing Dictator or Ultimatum games. In our experiment, robbers sent on average 34.44% of the available resources as trustors and returned 17.99% as trustees. These levels seem to be within the standard ranges in the literature as reported, e.g., in the meta-analysis of Johnson and Mislin (2011), although on the lower range of trust and (especially) trustworthiness.[12] The closest design we could find in a published study is Bellemare and Kröger (2007), which employed the strategy method in a laboratory Trust Game played with students in central Europe where transfers were doubled. In that study, trustors sent on average 30.58% of their resources, and trustees send back around 24% of what was available.[13] Our results are also close to those of Eckel and Petrie (2011).

The Big Robber Game and Social Preferences. In summary, results are strikingly consistent across all games. Robbers who took the maximum gave less in the Dictator Game, sent less money back in the Trust Game as trustees, made lower offers in the Ultimatum Game, trusted less as trustors in the Trust Game, and accepted lower offers as responders in the Ultimatum Game. In view of this evidence, it can be concluded that the Big Robber Game, as intended, measures a correlate of social preferences *as they have been discussed in the literature until now*, and not a different construct.

We remark that the fact that behavior is qualitatively consistent between the Big Robber Game and all other games is of independent interest, since such associations should not be taken for granted at the individual level. Blanco et al. (2011) estimated two parameters of inequality aversion in a within-subject design where participants played several games, including Dictator and Ultimatum games, and found remarkably low correlations within subjects. Galizzi and Navarro-Martínez (2018) found low correlations between behavior in Dictator, Ultimatum, and Trust games in the lab and social behavior in the field. This is in line, e.g., with Stoop et al. (2012), who found clear divergences in cooperation levels between lab and field settings. However, other authors have found positive associations across domains. Dariel and Nikiforakis (2014) found a qualitative correlation for prosocial behavior across games, with participants behaving cooperatively in a public-good game reciprocating higher wages with higher effort levels in a gift-exchange game. Fisman et al. (2007) find a strong positive association between preferences for giving and social preferences referred to distributions among others (which do not affect the own payoffs).

[12] In view of the literature, the fact that our results for the Trust Game are in the lower range might be unsurprising. We used a multiplying factor of two, while many studies use a factor of three. Johnson and Mislin (2011) suggest that a lower factor reduces trustors' transfers and trustees' returns. Casari and Cason (2009) argue that the strategy method, which we used, reduces transfers of trustees (although Brandts and Charness, 2011 find no difference). Burks et al. (2003) argue that playing both roles in the Trust Game reduces both overall trust and overall reciprocity; in our setting, players first made a trustor decision and were asked for a trustee decision later on.

[13] We thank Sabine Kröger for providing the aggregate statistics. In that study, however, trustees had an additional endowment.

4 Does the Big Robber Question Affect Behavior?

In this section, we focus on treatment effects. The Big Robber decision is clearly not one that can be taken lightly. It is conceivable that behavior in the Dictator, Ultimatum, and Trust games was affected by having answered this question beforehand. This is precisely the reasoning which brought us to include the ex ante and ex post treatments.

There are indeed a number of natural hypotheses grounded on psychological research. According to the "what-the-hell effect," an initial loss of self-control can lead to a modal change in which all pretenses are abandoned, for example in the dieting domain (Polivy and Herman, 1985). It has been argued (Achtziger et al., 2015, 2016, 2018) that prosocial behavior requires self-control in order to override selfish impulses. A natural implication then would be that, if a participant has revealed being selfish in the ex ante Big Robber decision, then he will behave less prosocially afterwards than controls (ex post robbers). However, the opposite hypothesis could also be justified. According to the "transgression-compliance effect," people who believe that they have harmed someone show an increased willingness to perform unrelated good deeds later on (Carlsmith and Gross, 1969), as if the latter could compensate the former (see Gneezy et al., 2014, for a recent illustration). This effect might reveal a mechanism to reduce experienced guilt. In our context, it would imply that if a participant has revealed to be selfish in the ex ante Big Robber decision, he or she should behave more prosocially than controls afterwards.

A further possible hypothesis concerns robbers who refrained from taking the maximum. According to the "moral credentialing" effect, humans often act as if an initial good behavior (even an exogenously induced one) provides a license to misbehave later on (Monin and Miller, 2001). For example, people purchasing "green" products are later on less likely to share in a Dictator Game, and more likely to cheat on a task to increase their gains (Mazar and Zhong, 2010; Zhong et al., 2010). According to moral credentialing, robbers who did *not* behave completely selfishly in the ex ante Big Robber decision should behave more selfishly afterwards than controls.

4.1 The more selfish robbers

We start with the behavior of robbers who took the maximum possible, i.e. robbers-50%, for which the opposed hypotheses derived from the what-the-hell effect and the transgression-compliance effect apply. In the Dictator Game, ex ante robbers-50% (N=100) gave on average 1.310 points, while ex post robbers-50% (N=80) gave on average 1.213 points. The difference is not statistically significant (MWW test, $z = 0.695$, $p = 0.4873$). However, looking at genders separately allows us to see a different picture. Figure 7 shows the relative frequency of points sent by the robbers-50% in the Dictator Game split by gender. In the ex ante treatment, female robbers-50% (N=49) sent on average 1.612 points while in the ex post treatment female robbers-50% (N=37) sent 0.811 points. This difference, which agrees with the possibly guilt-induced transgression-

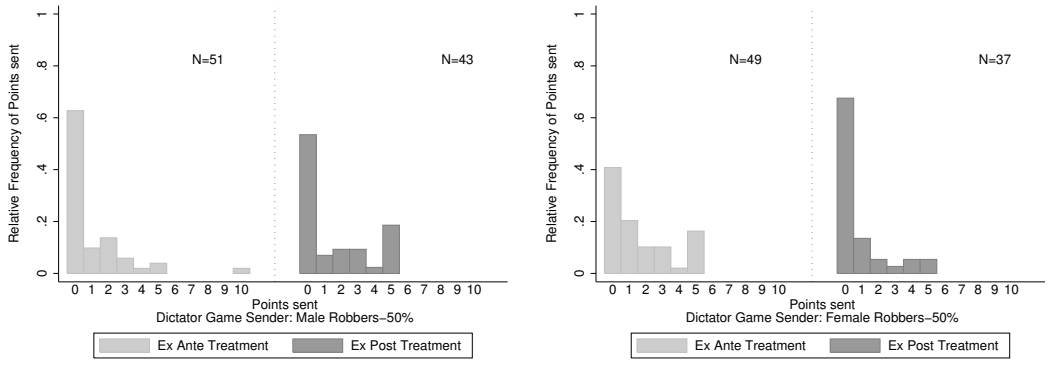

Figure 7: Behavior of Robbers-50% in the Dictator Game. Relative frequency of points sent by the robbers-50% in the Dictator Game split by treatment and gender (left-hand side: males; right-hand side: females).

compliance effect, is highly significant according to a Mann-Whitney-Wilcoxon test ($z = 2.442$, $p = 0.0146$). In contrast, looking at male robbers only, there is no evidence of the transgression-compliance effect. Ex ante male robbers-50% (N=51) sent on average 1.020 points in the Dictator Game, compared to 1.558 points sent by ex post male robbers-50% (N=43). The difference goes in the opposite direction, that is, the one predicted by the what-the-hell effect, but does not reach significance (MWW test, $z = -1.348$, $p = 0.1775$).[14]

Moving to the behavior of robbers-50% in the Ultimatum Game, offers in the ex ante treatment ($N = 100$, mean 3.760) were significantly lower than the offers of the robbers-50% in the ex post treatment ($N = 80$, mean 4.125 points; MWW test, $z = -1.808$, $p = 0.0705$). This is in agreement with the what-the-hell effect. However, this finding is driven by male behavior. Figure 8 shows the relative frequency of points proposed by the robbers-50% in the Ultimatum Game split by gender. Male robbers-50% in the ex ante treatment ($N = 51$) offered 3.765 points on average, which was significantly lower than the average 4.209 points offered by the male robbers-50% ($N = 43$) in the ex post treatment (MWW test, $z = -2.155$, $p = 0.0312$). In contrast, there were no significant differences for females (ex ante, $N = 49$, average 3.755; ex post, $N = 37$, average 4.027; MWW test, $z = -0.319$, $p = 0.7495$).

The decision in the Dictator Game and the proposer decision in the Ultimatum Game were the two first decisions in the sequence of games. For those, as commented above we find small effects which depend on gender. The responder decision in the Ultimatum game and the two decisions (as trustor and as trustee) in the Trust Game are the last three decisions in the sequence of games, and behavior might have been affected by the most recent decisions (in the Dictator and Ultimatum games) more than by the

[14]Out of $N = 640$ participants, only 3 gave 10 points in the Dictator Game, all of them male: one victim and two robbers. One of the robbers took 50%. Excluding this atypical individual from the test, the average points sent by ex ante male robbers-50% (N=50) drop to 0.840 points and the difference between treatments becomes more clear, but still misses significance (MWW, $z = -1.561$, $p = 0.1185$).

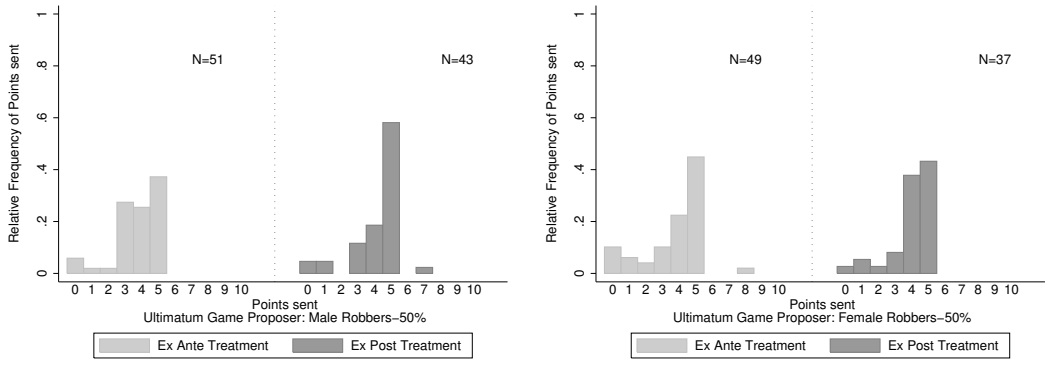

Figure 8: Behavior of Robbers-50% in the Ultimatum Game. Relative frequency of points proposed by the robbers-50% in the Ultimatum Game split by treatment and gender (left-hand side: males; right-hand side: females).

previous Big Robber question. Hence, we did not expect strong treatment differences for these decisions. Indeed, we find no significant differences, even when looking at genders separately, in Minimum Thresholds of Acceptance in the Ultimatum Game or in behavior in the Trust Game between ex ante and ex post robbers-50%.

We conclude that the psychological effects caused by the Big Robber decision on more selfish robbers might be present but are, first, generally weak, and, second, dependent on gender. While women seem to be affected by guilt-related considerations as the transgression-compliance effect, men seem to be show the opposite tendency as predicted by the what-the-hell effect.

4.2 The less selfish robbers

The behavior of robbers who declined to take the maximum possible, i.e. robbers-no-50%, might be affected by the moral-credentialing effect, that is, since they did not behave (maximally) selfishly in the Big Robber decision, they might have behaved more selfishly in the subsequent games. However, we find no evidence for such an effect in the Dictator Game. Ex ante robbers-no-50% ($N = 60$) gave an average of 3.200 points, which is not significantly different from the average of 3.125 given by ex post robbers-no-50% ($N = 80$) (MWW test, $z = -0.172$, $p = 0.8633$). The differences remain insignificant when splitting the sample by gender.

In the Ultimatum Game, there are also no differences for the full sample. Ex ante robbers-no-50% ($N = 60$) proposed an average of 4.500 points, which is not significantly different from the average of 4.588 proposed by ex post robbers-no-50% ($N = 80$) (MWW test, $z = -0.148$, $p = 0.8820$). However, splitting the sample by gender reveals (marginally) significant and opposed effects. Ex ante female robbers-no-50% ($N = 31$) offered an average of 4.323 points while ex post female robbers-no-50% ($N = 43$) offered an average of 4.767 points (MWW test, $z = -1.701$, $p = 0.0889$). This difference is in the direction predicted by the moral-credentialing effect. In contrast, the treatment dif-

ference for male robbers-no-50% went in the opposite direction, with significantly higher offers in the ex ante treatment ($N = 29$, mean 4.690) compared to the ex post treatment ($N = 37$, mean 4.378; MWW test, $z = 1.657$, $p = 0.0974$).

As in the case of robbers-50% (and most likely for identical reasons), we do not find any significant treatment differences for robbers-no-50% as responders in the Ultimatum Game or in either role in the Trust Game.

We conclude that, as in the case of more selfish robbers, the psychological effects caused by the Big Robber decision on less selfish robbers are generally weak and depend on gender. While women seem to be affected by the moral-credentialing effect in the Ultimatum Game, men show the opposite tendency.

4.3 The victims

All victims in the experiment ($N = 320$) learned about the possibility that part of their earnings could be taken away before making their decisions in the sequence of games. It is natural to ask whether this information affected their behavior. For instance, one could speculate that, knowing that part of their earnings might be lost, they might have become more self-centered. One could also speculate with possible negative reciprocity against the robbers, although this is an unlikely motivation because victims did not know whether the partners they would be matched with to play the Dictator, Ultimatum, and Trust games would be victims or robbers (random matching).

The cleanest comparison to find possible effects in victims' behaviors is with ex post robbers ($N = 160$). These robbers were informed about the Big Robber decision after they played the Dictator, Ultimatum, and Trust games, hence their behavior in those games could not be affected by the latter question, and they are in practice a control group. Comparing the behavior of victims to the behavior of ex post robbers, however, reveals no significant differences. In the Dictator Game, victims sent an average of 2.216 points, compared to 2.169 sent by ex post robbers (MWW test, $z = 0.217$, $p = 0.8282$). In the Ultimatum Game, victims proposed an average of 4.359 points, compared to 4.356 proposed by ex post robbers (MWW test, $z = -0.049$, $p = 0.9609$). As responders, the victims' average MTA ($N = 308$) was 2.912, compared to an average MTA of 2.891 for ex post robbers ($N = 156$; MWW test, $z = -0.042$, $p = 0.9662$). In the Trust Game, victims sent an average of 4.013 points, compared to 3.575 sent by ex post robbers (MWW test, $z = 1.420$, $p = 0.1556$). As trustees, victims sent back an average of 20.21%, compared to 18.75% sent back by ex post robbers (MWW test, $z = 0.797$, $p = 0.4252$). Splitting the tests by gender revealed only one significant effect, namely that male victims ($N = 146$) sent an average of 4.480 points while male robbers ($N = 80$) sent an average of 3.625 points, with the difference being significant according to a Mann-Whitney-Wilcoxon test ($z = 1.793$, $p = 0.0730$). However, none of the other nine within-gender comparisons for the five decisions was significant.

We conclude that victims' behavior was unaffected by the knowledge that a part of their earnings might be taken away by robbers.

5 Donations

Since the Big Robber decision is a high-stakes one, it is natural to ask whether, in some or even most of the cases, the decision to take the maximum might have led to feelings of guilt and regret. There is evidence from social psychology pointing out that guilt can motivate prosocial behavior perceived as reparative, as a way to appease guilty feelings (Malinowski and Smith, 1985; Giner-Sorolla, 2001; Zemack-Rugar et al., 2007). In particular, Lindsey (2005) showed that guilt is associated with increased charity donations. This is also consistent with Andreoni (1989), who argued that people derive direct utility from the act of giving to a charity, with Gneezy et al. (2014), who showed that people were then more likely to donate to charity after making an immoral choice, and with Andreoni et al. (2017), who showed that people give more if exposed to stimuli which activate empathy ("Please give!") but at the same time try to avoid exposure to such stimuli.[15]

We hence decided to include a final question giving participants the opportunity to donate part of their earnings to a charity, a local animal shelter.[16] In the final questionnaire, we also asked for the valuation of the charity organization on a scale from 1 = very bad to 10 = very good. The average evaluation was 6.46 ($N = 640$, SD= 2.66), indicating a generally positive view.

The form of the question was as follows. "You can donate a fraction of your total earnings (excluding the show-up fee) to a charity organization (name of animal shelter). How much do you want to donate?" The question was posed after all decisions had been made, but before the actual Big Robber of the session was selected and the game decisions were implemented. On average, subjects donated a fraction of 5.31% of their earnings, for a total joint donation of €299.6. Taken as a whole, that is, not differentiating among those who took the maximum and those who did not, the donations of robbers did not differ from those of the victims. Victims ($N = 320$) donated on average 6.10% while robbers (pooled, $N = 320$) donated on average 4.52%. The difference is not significant (MWW test, $z = -0.122$, $p = 0.9029$).

We hypothesized that, if feelings of guilt or regret were associated with the decision to take the maximum, robbers who took 50% would donate more than others. On the contrary, if the decision to take the maximum simply reflects a stronger tendency to rely on pure self-interest, we should observe lower donations. We found that robbers who took the maximum donated significantly less than others. Robbers-50% ($N = 180$) donated

[15] DellaVigna et al. (2012) pointed out that social pressure might be an additional motive for giving which needs to be dissociated from altruism. In our setting social pressure plays no role since all decisions are made anonymously. However the need to appease feelings of guilt might be seen as individual pressure coming from self-perception after facing the Big Robber decision.

[16] The charity was the "Cologne Animal Protection Club of 1868."

an average of only 2.62%, compared to 6.96% donated by robbers-no-50% ($N = 140$). The difference is both substantial and highly significant according to a Mann-Whitney-Wilcoxon test ($z = -5.934$, $p < 0.0001$) The effect remains for each treatment. Ex ante robbers-50% ($N = 100$) donated only 3.12%, while ex ante robbers-no-50% ($N = 60$) donated 6.20% (MWW test, $z = -3.339$, $p = 0.0008$). Ex post robbers-50% ($N = 80$) donated a mere 2.00%, while ex post robbers-no-50% ($N = 80$) donated 7.54% (MWW test, $z = -4.982$, $p < 0.0001$). The left-hand side of Figure 9 depicts the average donations of robbers-no-50% and robbers-50% for both treatments. The effect also persists when looking at each gender separately.[17]

A possible criticism of this test is that robbers who took 50% expected higher earnings than those who did not, hence they might have tried to adjust down the donation, expressed as a percentage, while still donating a larger absolute amount. This is not the case. We computed the expected donation under the extremely generous assumption that each individual robber might have assumed that he or she would indeed be the selected one (rather than using a probability of 1/16). Under this assumption, the expected donation of robbers-50% would have been €2.89, compared to an expected donation of robbers-no-50% of €3.80. The difference is again highly significant (MWW test, $z = -4.114$, $p < 0.0001$). That is, no matter how generous the criterion, robbers who took the maximum also donated less in absolute terms.[18] The right-hand side of Figure 9 shows the average donations in percentages (left axis) and the expected donations in absolute terms (right axis) for the different robber groups and the victims.

In conclusion, there is no evidence of guilt or regret as captured by the donation decision. One could of course speculate that the donation decision is only moral at an abstract level in the sense that it does not affect the victims. Also, one could invoke a different version of the what-the-hell effect to explain the results. The simplest explanation, however, is that robbers who took the maximum are just more selfish in all dimensions than those who did not.

It is reasonable to ask whether the differences in donations where caused by differences in evaluations, or whether they were rationalized ex post by providing lower evaluations of the charity. In both cases, we would expect robbers-50% to provide worse evaluations. This is, however, not true. Robbers-50% ($N = 180$) rated the charity on

[17] Pooling all robbers together, there was no significant difference in donations across treatments. There is a small gender difference in donations, with male robbers donating less (3.91%) than female robbers (5.13%). The difference is marginally significant (MWW test, $z = -1.651$, $p = 0.0988$). This gender difference is also found among the victims, with male victims ($N = 146$) donating an average of 3.29%, compared to an average of 8.46% by female victims ($N = 174$). The difference is highly significant (MWW, $z = -3.691$, $p = 0.0002$).

[18] However, robbers as a whole donated more in absolute terms than victims, which is not surprising since they simply earned substantially more on average. Average expected donations of the victims were €0.63, compared to €3.29 for the robbers. This difference is highly significant (MWW test, $z = -4.762$, $p < 0.0001$).

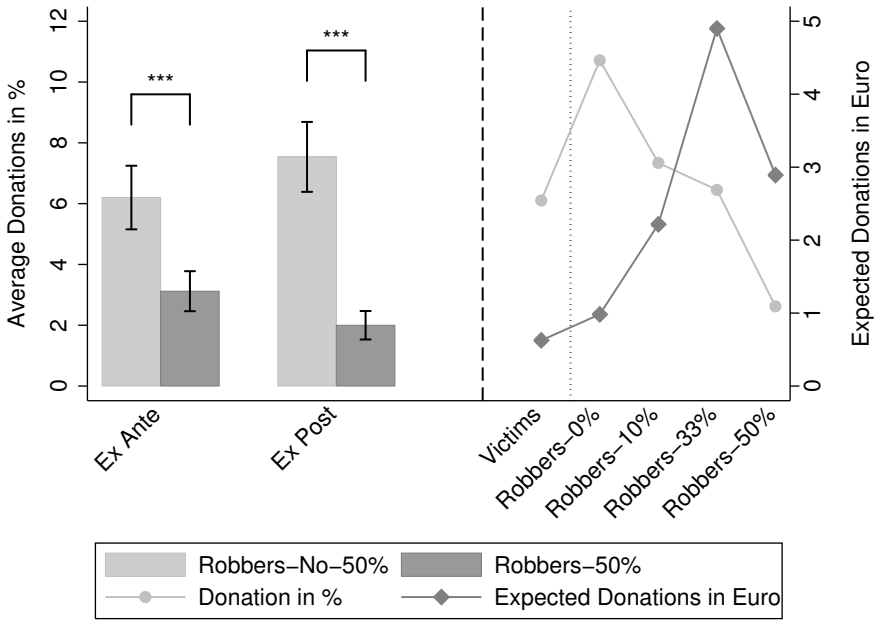

Figure 9: Average donations in percentage and expected donations in Euros. The left-hand side shows the donations of robbers-no-50% and robbers-50% in both treatments. The right-hand side shows the donations in percentage (left axis) and the expected donations in Euros (right axis) for each robbing group and victims. *** $p < 0.01$, MWW test.

average with 6.69, while robbers-no-50% (N=60) rated it lower, at 6.28. The difference goes in the opposite direction but is not significant (MWW test, $z = 1.506$, $p = 0.1320$).[19]

6 Models of Social Preferences

In this section, we examine whether the behavior we observe is compatible with received models of social preferences, focusing on the models of Fehr and Schmidt (1999) (hereafter FS), Bolton and Ockenfels (2000) (hereafter BO), Charness and Rabin (2002) (hereafter CR), and Alger and Weibull (2013) (hereafter AW). The strategy of analysis is a simple out-of-sample exercise as follows. Within the sequence of games, the very first decision of each participant corresponded to that of a dictator in the Dictator Game. From this decision, we deduce the individual parameters of the utility functions proposed in each of the models by FS, BO, CR, and AW. On the basis of these parameters, we derive the predicted behavior in the Big Robber Game for each participant and model. Finally, we compare predicted with actual behavior and examine the "fit" of the models,

[19]Interestingly, ex ante robbers-50% ($N = 100$) rated the charity on average with 6.91, while robbers-no-50% ($N = 60$) rated it at 6.13, which is significantly lower (MWW test, $z = 1.835$, $p = 0.0666$). Hence, ex ante robbers who took the maximum donated less in spite of the fact that they valued the charity better. In the ex post treatment the valuations of the robbers-50% ($N = 80$, mean 6.43) were not significantly different from the valuations of the robbers-no-50% ($N = 80$, mean 6.39; MWW, $z = -0.100$, $p = 0.9205$).

simply by examining the percentage of decisions in our sample of 320 robbers which are compatible with the model predictions.

Let $x_r \in [0, 10]$ be the amount sent by the dictator to the receiver. The model of CR reduces to the following parametric family of utility functions for the dictator.

$$U_D^{CR}(x_r) = \begin{cases} 10 - x_r - \rho \cdot (10 - 2x_r) & \text{if } x_r \leq 5 \\ 10 - x_r - \sigma \cdot (10 - 2x_r) & \text{if } x_r > 5 \end{cases}$$

where ρ and σ are parameters capturing distributional preferences (an additional parameter in the formulation of CR captures reciprocity considerations which play no role here).

This model encompasses the one of FS, with $\beta = \rho$ being the parameter for advantageous inequality and $\alpha = -\sigma$ the one for disadvantageous inequality. The model of FS further constrains $\sigma < 0 < \rho < 1$. In our setting, the only difference between both models is that with the additional constraints, FS' model cannot explain dictator decisions giving strictly more than 5 to the receiver; as a consequence, strictly speaking 5 out of our 320 observations would remain unclassified in the FS case. Otherwise, the analysis for FS and CR is identical.

In our sample, 123 (38.44%) of the robbers gave 0 in the Dictator Game, which implies $\rho \leq 1/2$ and $\sigma \leq 1 - \rho$; 130 (40.63%) robbers gave between 1 and 4 in the Dictator Game, which implies $\rho = 1/2$ and $\sigma \leq 1/2$; 62 (19.38%) robbers gave exactly 5, which implies $\rho \geq 1/2$ and $\sigma \leq 1/2$; 3 (0.94%)robbers gave between 6 and 9, which implies $\rho \geq 1/2$ and $\sigma = 1/2$; and 2 (0.63%) robbers gave exactly 10, implying $\sigma \geq 1/2$ and either $\rho \geq 1/2$ or $\rho < 1/2$ together with $\sigma \geq 1 - \rho$. The first three columns of Table 2 summarize these observations.

Let $E[\Pi_V]$ denote the expected income of all victims in a session from the intermediate games (from the point of view of the individual robber) and $E[\Pi_R]$ the own (robber) expected income from those games. Let n_V be the number of victims, and denote by p the share the robber chooses to rob. A given robber would expect earnings $E[\Pi_R] + pE[\Pi_V]$, and will expect average earnings of the victims to be $(1 - p)(1/n_V)E[\Pi_V]$. For the Big Robber, the CR (or FS) model implies the following utility function.

$$\begin{aligned} U_R^{CR}(p) &= (E[\Pi_R] + pE[\Pi_V]) - \rho \left(E[\Pi_R] + pE[\Pi_V] - (1 - p)\frac{1}{n_V}E[\Pi_V] \right) \\ &= (1 - \rho)E[\Pi_R] + \frac{1}{n_V}E[\Pi_V]\left(p \cdot (n_V - \rho(n_V + 1)) + \rho\right). \end{aligned}$$

Note that the parameter σ plays no role because, in expected terms, the robber could not be worse off than the victims.

Denote by V the ex ante expected earnings of all n_V victims in a session as communicated in the experiment. For an ex ante robber, since the decision to rob or not was made before the intermediate games were explained and played, all the player could

deduce was $E[\Pi_V] = V$ and $E[\Pi_R] = V/n_V$. Hence, for an ex ante robber, the expression above simplifies to

$$U_R^{CR}(p) = \frac{1}{n_V} V \left[1 + p \cdot (n_V - \rho(n_V + 1))\right].$$

For an ex post robber, however, the intermediate games had already been played when the Big-Robber decision was made, and expectations could have been adjusted. For instance, those robbers knew that they would be receiving at least $20 - x_r - x_r^T$, where x_r was their decision in the Dictator Game and x_r^T the decision as a trustor in the Trust Game, and accordingly could adjust $E[\Pi_V]$ down. However, the optimum of $U_R^{CR}(p)$ is independent of the exact values of expectations, and in particular there should be no difference between ex ante and ex post robbers according to the CR model. For all $\rho < \frac{n_V}{n_V+1} = \frac{16}{17}$ robbers should take 50%, while for $\rho = \frac{16}{17}$ robbers are indifferent among taking 50%, 33%, 10%, or 0%, and for $\rho > \frac{16}{17}$ robbers should take 0%. Hence, all robbers who gave strictly less than 5 points in the Dictator Game should take 50%, and we have no prediction for those who gave exactly 5 points or more than 5 points, in the sense that for $\rho > 1/2$, their optimum might be to take 50%, 0%, or correspond to full indifference among all possibilities. The second to last column of Table 2 summarizes these predictions and the amount of observations compatible with these predictions.

Because of the linearity of their formulation, the models of FS and CR can only explain intermediate values (1 to 4 points) in the Dictator Game through indifference. The model of BO incorporates a nonlinear term to deal with this and related issues, the typical implementation relying on a quadratic functional form. Applied to the Dictator Game, the corresponding utility function is as follows.

$$U_D^{BO}(x_r) = a(10 - x_r) - \frac{b}{2}\left(\frac{10 - x_r}{10} - \frac{1}{2}\right)^2 = a(10 - x_r) - \frac{b}{2}\left(\frac{5 - x_r}{10}\right)^2$$

where again x_r is the amount the dictator sends to the receiver, and a and b are parameters weighting the utility of the own payoff and disutility of the relative payoff, for which BO assume only $a \geq 0$ and $b > 0$. A player's type is fully characterized by the ratio a/b. The fourth column of Table 2 shows the ranges of a/b implied by the observed behavior of robbers in the Dictator Game.[20]

For the Big Robber, the BO model implies the following utility function.

$$(1) \qquad U_R^{BO}(p) = a\left(E[\Pi_R] + pE[\Pi_V]\right) - \frac{b}{2}\left(\frac{E[\Pi_R] + pE[\Pi_V]}{E[\Pi_R] + E[\Pi_V]} - \frac{1}{n_V + 1}\right)^2.$$

[20]Each possible decision implies a different range, but all decisions in the $1-4$ range, and analogously for the $6-10$ range, yield the same predicted behavior in the Big Robber Game, hence we do not differentiate them in the table.

Again, for an ex ante robber $E[\Pi_V] = V$ and $E[\Pi_R] = V/n_V$, hence the expression above simplifies to

$$U_R^{BO}(p) = a\frac{V}{n_V}(1 + p \cdot n_V) - \frac{b}{2}\left(\frac{n_V}{n_V + 1}p\right)^2.$$

Since experimental sessions involved 32 participants, half of which were victims, we have that $n_V = 16$. The expected income of all victims was €200, corresponding to 571.43 points (recall that the Dictator Game decisions was how to split 10 points). With these values, a direct computation shows that $U_R^{BO}(p)$ for ex ante robbers is maximized at $p = 0.5$ or above for any ratio a/b larger than 0.00064. In view of the values derived from behavior in the Dictator Game, all ex ante robbers who gave less than 5 points in the Dictator Game are predicted to take 50%, those who gave 6 points or more are inconsistent with the model, and for those who opted for an equal split, the possible values derived for a/b do not allow a precise prediction (hence their behavior in the Big Robber Game is always aligned with the model).

Ex post robbers form their expectations about their own payoff from the intermediate games depending on their actions. In our design, the robbers play against a different player in each intermediate game but do not know if the opponent is a robber or victim. If the robbers played only against robbers, the expectations about the victims earnings should be unaffected, $E[\Pi_V] = V$. In the opposite extreme, all six interactions of the robber in the intermediate games might have affected victims. It follows that on average, at most one victim (playing in six games) was affected by the robber decisions, implying that $E[\Pi_V] \in [(n_V - 1)V/n_V, V]$. From equation (1) above, we obtain that the (unconstrained) maximum is reached at

$$p^* = \frac{a}{b}\left(\frac{1}{E[\Pi_V]}\left(E[\Pi_R] + E[\Pi_V]\right)^2\right) + \frac{1}{n_V + 1}\left(\frac{E[\Pi_R]}{E[\Pi_V]} + 1\right) - \frac{E[\Pi_R]}{E[\Pi_V]}$$

and the derivative of this expression with respect to $E[\Pi_V]$ is

$$\frac{\partial p^*}{\partial E[\Pi_V]} = \frac{a}{b}\left(1 - \left(\frac{E[\Pi_R]}{E[\Pi_V]}\right)^2\right) + \frac{n_V}{n_V + 1}\frac{E[\Pi_R]}{E[\Pi_V]^2}.$$

We know that $E[\Pi_R] \in [20 - x_r - x_r^T, 70 + x_r^T]$, where the maximum possible payoff of 80 points results if the robber gave the minimum away except for the Trust Game where she gave everything to the trustee and receives the maximum possible in each of the intermediate games. Since $V = 571.43$ points and $n_V = 16$, no matter what the exact expectation adjustment was, $E[\Pi_R] < (n_V - 1)V/n_V \leq E[\Pi_V]$. It follows that the derivative of p^* with respect to $E[\Pi_V]$ is strictly positive, which implies that, in the BO model, robbers should rob (weakly) more with larger expectations on $E[\Pi_V]$.

Analogously, the derivative of the expression of p^* with respect to $E[\Pi_R]$ is

$$\frac{\partial p^*}{\partial E[\Pi_R]} = \frac{a}{b}\frac{2}{E[\Pi_V]}\left(E[\Pi_R] + E[\Pi_V]\right) - \frac{n_V}{n_V + 1}\frac{1}{E[\Pi_V]}$$

and is larger than zero for all ratios a/b above $\frac{n_V}{n_V+1}\frac{1}{2(E[\Pi_R]+E[\Pi_V])}$. The latter expression cannot be larger than its value for the minimum feasible values of $E[\Pi_R]$ and $E[\Pi_V]$, which are 0 and $\frac{15}{16}571.43$ points, respectively. With $n_V = 16$, this implies that p^* is increasing in $E[\Pi_R]$ for all a/b above (approximately) 0.00088. That is, for such values, in the BO model, robbers should also rob (weakly) more with larger expectations on $E[\Pi_R]$. Thus, the optimum p^* will be larger than the value of p^* with the smallest possible expectation $E[\Pi_V] = (n_V - 1)V/n_V$ and the lower bound $E[\Pi_R] = 0$, that is,

$$\frac{a}{b}\frac{n_V - 1}{n_V}V + \frac{1}{n_V + 1}$$

but with $n_V = 16$ and $V = 571.43$, this expression is always larger than 0.5 for any a/b above (approximately) 0.00082. In view of the values derived from behavior in the Dictator Game, all ex post robbers who gave less than 5 points in the Dictator Game are predicted to take 50%, those who gave 6 points or more are inconsistent with the model, and only for those who opted for an equal split, the possible values derived for a/b do not allow a precise prediction. That is, the model predicts no difference between ex ante and ex post robbers. In particular, the implications are in practice identical to those derived from the CR model.

The model of AW states that an individual is a homo moralis if her utility function is the convex combination of selfishness (maximization of own payoff) and Kantian morality (the payoff received when everybody acts as the individual). The weight each subject puts on the moral payoff reveals her degree of morality, i.e., $\kappa \in [0, 1]$ with $\kappa = 0$ being completely selfish (homo oeconomicus) and $\kappa = 1$ being completely moral (or homo kantiensis).

The AW model considers symmetric games, but it is extended to asymmetric ones as the Dictator Game by recasting it as a role game, that is, the player considers herself as either the dictator or the receiver with probability $\frac{1}{2}$ each (Alger and Weibull, 2013, Section 6.1.1). This yields the utility function

$$U_D^{AW}(x_r) = \frac{1}{2}\left(v\left(\frac{10 - x_r}{10}\right) + \kappa \cdot v\left(\frac{x_r}{10}\right) + (1 - \kappa) \cdot v\left(\frac{y_r}{10}\right)\right)$$

where again x_r is the amount the dictator sends to the receiver, y_r is the amount received if the agent is the receiver, and $v : [0, 1] \mapsto \mathbb{R}$ is a differentiable function with $v' > 0$ and $v'' < 0$ representing the well-being from wealth.

Hence, given a function v we can infer bounds on the degree of morality κ that subjects reveal depending on what they sent in the Dictator Game. For our exercise we define $v = \sqrt{x}$ which satisfies the required properties for v stated above.[21] The ranges

[21] The model of BO is also cast for general functional forms, but the functional form we use is the one typically employed in the literature. Hence, although using a specific functional form for the model of AW is arbitrary, it keeps the exercise comparable with the previous one.

for the implied κ are displayed in the fifth column in Table 2. Note that, as in the case of BO, those who gave 6 points or more are inconsistent with the AW model.

For the Big Robber, the AW model (with a $\frac{1}{2}$ probability of being a robber or a victim) implies the following utility function.

$$U_R^{AW}(p) = \frac{1}{2} \cdot \left(v\left(\frac{E[\Pi_R] + p \cdot E[\Pi_V]}{E[\Pi_R] + E[\Pi_V]}\right) + \kappa \cdot v\left(\frac{(1-p) \cdot E[\Pi_R]}{E[\Pi_R] + E[\Pi_V]}\right) \right. \\ \left. + (1-\kappa) \cdot v\left(\frac{(1-q) \cdot E[\Pi_R]}{E[\Pi_R] + E[\Pi_V]}\right) \right) \tag{2}$$

where p is as defined above, q is the share that is taken from the participant in case she is a victim, and v is defined as in the Dictator Game above. Note that the the second and third terms refer to hypothetical situations where the decision maker becomes a victim. However, the decision maker still has the same expectation $E[\Pi_R]$ on her earnings from the intermediate games, and the term $E[\Pi_R] + E[\Pi_V]$ should be interpreted as the sum of this expectation and the expected earnings by all other players in those games.

Once again, for an ex ante robber $E[\Pi_V] = V$ and $E[\Pi_R] = V/n_V$, hence the expression above simplifies to

$$U_R^{AW}(p) = \frac{1}{2}\left(v\left(\frac{1 + p \cdot n_V}{1 + n_V}\right) + \kappa \cdot v\left(\frac{1-p}{1+n_V}\right) + (1-\kappa) \cdot v\left(\frac{1-q}{1+n_V}\right)\right)$$

A direct computation shows that, for ex ante robbers, $U_R^{AW}(p)$ is maximized at $p = 0.5$ or larger for any level of morality $\kappa \in [0, 1]$. Therefore we conclude that all ex ante robbers with $\kappa \in [0, 1]$ should take 50%.

Regarding ex post robbers, we proceed as in the BO model. The optimal p^* that maximizes the subject's utility given by (2) is

$$p^* = \frac{E[\Pi_V]^2 - \kappa^2 E[\Pi_R]^2}{E[\Pi_V]E[\Pi_R]\kappa^2 + E[\Pi_V]^2}$$

The derivative of this expression with respect to $E[\Pi_V]$ is

$$\frac{\partial p^*}{\partial E[\Pi_V]} = \frac{E[\Pi_R]\kappa^2\left(E[\Pi_V]^2 + E[\Pi_R]^2\kappa^2 + 2E[\Pi_V]E[\Pi_R]\right)}{\left(E[\Pi_V]E[\Pi_R]\kappa^2 + E[\Pi_V]^2\right)^2}$$

which is strictly positive for $\kappa, E[\Pi_V], E[\Pi_R] > 0$. That is, in the AW model, robbers should rob (weakly) more with larger expectations on $E[\Pi_V]$. Note that completely selfish agents, i.e. those $\kappa = 0$, will of course rob as much as possible.

Analogously, the derivative of the expression of p^* with the respect to $E[\Pi_R]$

$$\frac{\partial p^*}{\partial E[\Pi_R]} = \frac{-E[\Pi_V]\kappa^2\left(\kappa^2 E[\Pi_R]^2 + 2E[\Pi_V]E[\Pi_R] + E[\Pi_V]^2\right)}{\left(E[\Pi_V]E[\Pi_R]\kappa^2 + E[\Pi_V]^2\right)^2}$$

Table 2: Out-of-sample analysis for the models of Charness and Rabin (2002), Bolton and Ockenfels (2000) and Alger and Weibull (2013).

x_r	Nr. Obs	Implied ρ (CR)	Implied a/b (BO)	Implied κ (AW)	Prediction BR Game	Compatible w/ prediction
0	123	$\leq \frac{1}{2}$	$[0.045, \infty[$	$[0, 0.162]$	Take 50%	100 (81.3%)
1-4	130	$= \frac{1}{2}$	$[0.005, 0.045]$	$[0.162, 0.904]$	Take 50%	59 (45.4%)
5	62	$\geq \frac{1}{2}$	$[0, 0.005]$		Undeterm.	62 (100%)
				$[0.904, 1]$	Take 50%	20 (32.3%)
6-9	3	$\geq \frac{1}{2}$	N/A	N/A	Undeterm.	3 (100%)
10	2	Undeterm.	N/A	N/A	Undeterm.	2 (100%)
Total	320				CR	226 (70.6%)
					FS/BO	221 (69.1%)
					AW	179 (55.9%)

is strictly negative for $\kappa, E[\Pi_V], E[\Pi_R] > 0$, which implies that robbers should rob (weakly) more with smaller expectations on $E[\Pi_R]$.

Thus, the optimum p^* will be larger than the value of p^* with the smallest possible expectation $E[\Pi_V] = (n_V - 1)V/n_V$ and the largest possible $E[\Pi_R] = 80$. Straightforward but cumbersome computations show that the corresponding value of p^* with $n_V = 16$ and $V = 571.43$ is always larger than 0.5 for any value of $\kappa \in [0, 1]$. This implies that ex post robbers with any level of morality should also take 50% in the Big Robber Game.[22] In summary, the AW model predicts the same as the other three models for robbers who gave 4 or less in the Dictator Game, but, while the models of FS, BO, and CR do not provide unique predictions for the 62 robbers who gave 5 in the Dictator Game, the AW model predicts that those robbers should take as much as possible, i.e. 50%. In our sample only 20 participants (32.26%) out of those 62 behaved as predicted by AW.

The last column of Table 2 compares the predictions of the models to actual behavior in the Big Robber Game. We obtain that 226 out of our 320 observations (70.63%) can be explained by the model of CR. Further, of the 226 observations, 159 (49.69% of the total, or 70.35% of the explained observations) are such that the robbers took the maximum, while 67 (20.94% of the total, or 29.65% of the explained) are such that the prediction of the model cannot be actually derived (because the participant gave 5 or more points in the Dictator Game). The 5 observations of robbers who gave strictly more than 5 points are inconsistent with the models of FS, BO, and AW. We obtain that 221 out of 320 observations (69.06%) can be explained by the models of FS and BR.

[22]This apparently surprising result has a straightforward intuition. The utility of an AW agent is a convex combination of selfishness and the hypothetical payoff obtained when everybody robs as much as the agent. Selfishness prescribes to rob as much as possible. The other part of the utility combines the large payoff increase obtained when the agent is a robber and robs p from 16 other agents with the comparatively small payoff decrease obtained when the agent is a victim and p is robbed from her. Obviously, the payoff increase dominates.

Out of the 221 observations, 159 are such that the robbers took the maximum, while 62 (19.38% of the total, or 28.05% of the explained) are such that the prediction of the model cannot be actually derived (because the participant gave exactly 5 in the Dictator Game). Only the AW model provides a unique prediction for subjects who gave exactly 5 in the Dictator Game. The AW model explains 179 out of 320 observations (55.94%), all of them such that the robbers took the maximum.

In summary, the observations are in line with received models of social preferences as FS, CR, or BO, although a significant percentage of observations are compatible with those models simply because they do not make unique predictions for participants who gave exactly 5 in the dictator game. It should be noted that, as indicated in Table 2, the standard models of social preferences we consider predict a larger number of participants taking 50% than we actually observe, and are hence even closer to selfish behavior than the data. The reason is that the very high potential payoff in the Big Robber Game, compared to the Dictator Game, makes it difficult to compare the ranges of the parameters across games. Therefore, each model comes to the conclusion that the robber should take the maximum of 50% whenever she gives less than the equal split in the Dictator Game. This coincides with the majority of our data. The model of AW fares considerably worse, essentially because it predicts that *homo moralis* will always rob as much as possible, which includes even those participants who gave 5 in the Dictator Game.

7 Discussion

The Big Robber Game is a paradigm which makes high stakes salient, while also emphasizing that the return comes at the cost of actively harming a large group of people. Hence, the paradigm directly captures the idea that a monetary temptation might make people act against society's interests, as in the many corporate scandals which have sparked public outrage in the recent decades, where "individuals who hold such financial power may be tempted to abuse it for their own personal profit" (Myerson, 2012, p.848). With this paradigm we showed that such corporate scandals are easily reproduced in the lab even with regular university students. An absolute majority of our (large) sample took the maximum possible amount, accepting that their decision would damage a large number of other people. Further, the decision to take the maximum was faster on average than the decision to refrain from it, revealing a weaker moral struggle in the former case, and those who took the maximum also donated less to a charity afterwards, revealing no evidence of guilt. Our results stand in sharp contrast with the hypotheses of Mazar et al. (2008), who argue that most people will cheat only a little if given opportunity.

At the same time, we show that the very same participants who are willing to inflict considerable damage on their peers within the Big Robber Game display standard levels of other-regarding behavior as reflected by Dictator, Ultimatum, and Trust games. That is, we empirically demonstrate the coexistence of prosociality in the small and morally

outrageous selfishness in the large, that is, in high-stakes, high-impact decisions affecting large groups.

The behavior we observe, however, is fully compatible with received models of social preferences. First, by calibrating standard models using behavior in the Dictator Game, we show that a large part of our data is consistent with the functional forms commonly-used to capture other-regarding preferences. If anything, those models predict higher levels of selfish behavior in the Big Robber Game than we actually observe. Second, we show that individual behavior in the Big Robber Game stands in a monotonic relation with behavior in the standard games mentioned above. Participants who took the maximum in the Big Robber Game gave less in the Dictator Game, offered less as proposers and accepted more unfair offers in the Ultimatum Game, and transferred less as trustors and returned less as trustees in the Trust Game. That is, behavior in all games, small and big, can be explained within a single account of other-regarding preferences. This is interesting in itself, in view of previous results on the (in)consistency of social preferences across games (e.g., Blanco et al., 2011).

Our experiment also allowed us to acquire a number of other insights. First, a number of psychological theories predict an effect of previous decisions in the moral domain on subsequent ones. We do find effects as predicted by those theories, but they are small and not systematic (that is, we find them for some games and not for others), and depend on gender. For men, we find evidence that behaving selfishly leads to further selfish behavior down the road, in accordance with the "what-the-hell" effect. For women, behaving selfishly seems to lead to a possibly guilt-induced attempt to behave less selfishly later on, in accordance with the "transgression-compliance" effect. Symmetrically, for women, refraining to behave selfishly earlier might lead to increasingly selfish behavior later, in accordance with the "moral-credentialing" effect. For men, however, refraining to behave selfishly earlier might lead to less selfish behavior later. Hence, while men seem to behave consistently, sticking to being more or less selfish, women act as if decisions added up and one bad canceled one good, and vice versa. We insist, however, that these effects are small and not systematic in our sample.

It is also noteworthy that we find no gender effect whatsoever in the Big Robber decision. Research from psychology suggests that women are more sensitive to social cues and feedback than men (Gilligan, 1982; Roberts and Nolen-Hoeksema, 1989). In line with this, several previous studies have identified gender differences in the degree of inequality aversion (for example, the proportion of egalitarian allocations in the Dictator and Ultimatum games). However, as shown in the review of Croson and Gneezy (2009), the evidence is mixed, and not all experiments find a clear gender difference. In a recent field experiment, Azmat et al. (2016) show that when stakes are high (as in our experiment), women perform worse than men, while the opposite is true whenever the stakes are low. This difference, though, refers to performance in real-effort tasks, while our experiment is about preferential choice.

The most important differences between the Big Robber Game and other games used to capture social preferences are, first, the fact that a single decision can inflict significant damage on a large group of people, and second, the size of the stakes. There are, of course, by necessity, a number of other differences between the Big Robber Game and other games. In our design, only one robber is selected out of 16 possible ones, and the individual decision is taken before the actual robber is selected. Hence there might be, in a certain sense, a diluted responsibility. However, this interpretation is at odds with evidence from the strategy method (Brandts and Charness, 2011), which shows that behavior is essentially aligned if decisions are made conditional on reaching a certain decision node and if the response is chosen after learning that the node has been reached. Another difference is that, in spite of the neutral framing, the decision in the Big Robber Game is more immediately placed into a moral framework than those in previous games, since it is not in terms of distribution or bargaining but simply in terms of taking away the earnings of others. The Big Robber Game is, simply put, a new paradigm, and not a high-stakes version of previous ones.

Since we aimed for a relatively large experiment in order to be able to test for treatment effects and condition those on gender, we settled on a single basic design rather than complicating the analysis with added variants from the onset. A large number of avenues for future research, involving design variants, are obvious at this point. The upper bound of 50% on what could be taken away from the victims was dictated by practical considerations, in order to avoid damaging the reputation of the lab by having half of the participants walk out of the experiment empty-handed. Future design variants could find ways to relax this constraint. Also, the discretization in only four possibilities reflected the desire to have clear behavioral groups (take nothing, take just a bit, take a significant part but not the maximum, and take the maximum). The baseline result having been thus established, it might be desirable to implement a continuous version. These and many other possibilities are beyond the scope of this paper and are left for future research.

In conclusion, the Big Robber Game contributes to the literature in three ways. From the methodological point of view, it provides the possibility to experiment with relatively high stakes without increasing existing research budgets. From the conceptual point of view, it provides a novel empirical demonstration that behavior which is commonly accepted as prosocial in "small" situations, as captured by standard laboratory games, coexists with behavior which is commonly considered morally questionable in "large" situations (high stakes, high impact). Last, the results echo popular discussions on moral responsibility among economic decision makers by showing that a large part of our sample (containing just regular university students) is willing to inflict significant damage on a relatively large number of people for personal gain, as long as that gain is of sufficient magnitude. Regrettably, our data suggests that, in our Western societies, hundred bucks might do the job.

Appendix: Translated Instructions

All instructions were on the screen. Original instructions were in German. Titles were generic ("part 1", etc). Text in [. . .] below contains clarifications, e.g. marking sentences shown only to some participants.

General Instructions

The experiment consists of two different parts and a questionnaire. The decisions you will make influence the payoffs you will receive at the end of the experiment.

Along the experiment you will earn ECUs (Experimental Currency Units). At the end of the experiment, your ECUs will be added up and exchanged into Euros. You will receive 35 Eurocents for 1 ECU. Independently of your decision, you will additionally receive 2.50 EUR for your participation in the experiment.

Instructions for the Big Robber Game

[*Text in this subsection was shown only to ex ante Robbers and Victims.*]

In this experiment, there are two types of participants. One half of the participants is Type I, the other half is Type II. [*Ex ante Robbers*:] You are Type I. [*Victims*:] You are Type II.

During the experiment, each participant will earn ECUs, that will be exchanged into Euros at the end of the experiment. The exact amount depends on your decisions and the decisions made by other participants during the situations described in the following instructions.

On the basis of the average earnings in previous experiments, we estimate that all Type II participants together will earn around 200 Euro. The maximum possible payment for all Type II participants together is 224 Euro.

At the end of this experiment, one of the Type I participants will be chosen at random to be the unique "receiver." This participant can decide which share of the earnings of all Type II participants he wants to have in addition to his own earnings. The earnings of the Type II participants will then be reduced by the corresponding amount.

Please note that all decisions and payments are made strictly anonymously. No other participants will know who the unique "receiver" was.

[*Ex ante Robbers*:] You now have to decide which share would like to receive from all Type II participants, in case you are the unique "receiver." The earnings of Type II participants will be reduced by the corresponding amount. The decision you make now will be implemented at the end of the experiment in case you are chosen to be the unique "receiver."

[*Victims*:] The possible shares that the unique "receiver" can choose from are as follows.

[*Table 1 was shown here.*]

Instructions for the Intermediate Games

You will now participate in six different decision situations.

In each decision situation, you will be paired with another participant. Depending on the decision situation, either you, the other participant, or both of you will make a decision. Your payment depends on these decisions. In every decision situation, you will be paired with a different participant.

In every decision situation, there are two different roles, participant A and participant B. Each decision situation will be explained to you in detail on the screen. Your role might change from one situation to the next. This will be displayed on the screen.

[*Dictator Game, Dictator*:] In this decision situation, there are two roles, participant A and participant B. Participant A receives 10 ECU and can keep them or send any part of it (any integer from 0 to 10) to participant B. The decision of participant A is binding and determines the final payments of both participants for this decision situation.

In this decision situation you are participant A. How many ECUs do you wish to send to participant B?

I send ... ECUs to participant B and keep the rest.

[*Ultimatum Game, Proposer*:] In this decision situation, there are two roles, participant A and participant B. A total of 10 ECUs are available. The participants decide together how to share the 10 ECU.

Participant A proposes a division of the 10 ECU to participant B. That is, participant A proposes how many ECUs participant B should receive and how many participant A keeps (the rest). After participant A proposes a division, participant B decides whether to accept or to reject the proposed division. If participant B accepts, the division is made exactly as participant A proposed. If participant B rejects the proposal, each of the participants receives 0 ECU.

In this decision situation you are participant A. How many ECUs do you offer to participant B?

My offer regarding the division of the 10 ECU is as follows. I propose that participant B receives ... ECUs and I keep the rest.

[*Trust Game, Trustor*:] In this decision situation, there are two roles, participant A and participant B. Participant A receives 10 ECU and can keep them or send any part of it (any integer from 0 to 10) to participant B. Each ECU sent is doubled. That is, for each ECU that participant A sends, participant B receives two ECUs. Afterwards, participant B decides how many ECUs he sends back to participant A and how many ECUs he keeps.

In this decision making situation you are participant A. How many ECUs do you wish to send to participant B?

I send ... ECUs to participant B and keep the rest.

Click on "calculate" to see how many ECUs you keep and how many ECUs participant B receives.

Click first on "calculate." Afterwards you will be able to confirm your choice.

Participant B receives double the amount that you send.

Participant B receives ... You keep ...

Additionally, you will receive the amount of ECUs that participant B sends back to you.

[*Dictator Game, Receiver*:] In this decision situation, there are two roles, participant A and participant B. Participant A receives 10 ECU and can keep them or send any part of it (any integer from 0 to 10) to participant B. The decision of participant A is binding and determines the final payments of both participants for this decision situation.

In this decision situation you are participant B. You do not have any decision to make. You will be informed about the decision of participant A at the end of the experiment.

[*Ultimatum Game, Responder*:] In this decision situation, there are two roles, participant A and participant B. A total of 10 ECUs are available. The participants decide together how to share the 10 ECU.

Participant A proposes a division of the 10 ECU to participant B. That is, participant A proposes how many ECUs participant B should receive (any integer from 0 to 10) and how many participant A keeps (the rest). After participant A proposes a division, participant B decides whether to accept or to reject the proposed division. If participant B accepts, the division is made exactly as participant A proposed. If participant B rejects the proposal, each of the participants receives 0 ECU.

In this decision situation you are participant B. For each possible offer of participant A, please decide whether you would accept or reject it.

[*Possible offers from 10 to 0 presented in list format.*]

[*Trust Game, Trustee*:] In this decision situation, there are two roles, participant A and participant B. Participant A receives 10 ECU and can keep them or send any part of it (any integer from 0 to 10) to participant B. Each ECU sent is doubled. That is, for each ECU that participant A sends, participant B receives two ECUs. Afterwards, participant B decides how many ECUs he sends back to participant A and how many ECUs he keeps.

In this decision situation you are participant B. Please indicate how many ECU you send back, depending on how many ECUs participant A sends.

[*Possible transfers from 10 to 1 presented in list format.*]

Instructions for the Big Robber Game

[*Text in this subsection was shown only to ex post Robbers. Instructions were identical to those for ex ante robbers again, except that instead of "the situations described in the following instructions," the instructions stated "the situations described in the previous instructions."*]

Instructions for Belief Elicitation

Consider the decision situation of the unique "receiver" in this experiment. Imagine that one participant of Type I in this experiment can decide to receive a share of the earnings of all Type II participants on top of his own payment. In total, Type II participants can earn up to 224 Euro. This means that a fraction of the earnings of all Type II participants will be transferred to this participant, depending on his or her decision. The earnings of Type II participants will be correspondingly reduced.

Out of 100 Type I participants, in your opinion, how many participants will choose each of the following options?

[*Big Robber options presented as above, as a list.*]

Please note that the total number of people must add up to 100. By clicking "calculate sum," you can see how many people you have already distributed.

Instructions for the Donation Question

You have the chance to donate a share of your payment to a charity. Your donation will support the (Name of the charity, an animal shelter) located at (Address of the charity).

The donation is a fraction of the sum that you will receive at the end of the experiment for all the decision situations, after the unique "receiver" has been selected and his or her decision has been implemented (it does not include your payment of 2.50 Euro for showing up for this experiment).

Which proportion of the amount mentioned above would you like to donate to (Name of the charity)?

Amount as a percentage (%) ...

How do you rate animal shelters? (1=very bad, ..., 10=very good)

References

Abbink, K., Brandts, J., Herrmann, B., and Orzen, H. (2010). Intergroup Conflict and Intra-group Punishment in an Experimental Contest Game. *American Economic Review*, 100(1):420–447.

Abbink, K., Brandts, J., Herrmann, B., and Orzen, H. (2012). Parochial Altruism in Inter-group Conflicts. *Economics Letters*, 117(1):45–48.

Abbink, K. and Herrmann, B. (2011). The Moral Costs of Nastiness. *Economic Inquiry*, 49(2):631–633.

Abbink, K., Irlenbusch, B., and Renner, E. (2000). The Moonlighting Game: An Experimental Study on Reciprocity and Retribution. *Journal of Economic Behavior and Organization*, 42(2):265–277.

Abbink, K. and Sadrieh, A. (2009). The Pleasure of Being Nasty. *Economics Letters*, 105(3):306–308.

Achtziger, A. and Alós-Ferrer, C. (2014). Fast or Rational? A Response-Times Study of Bayesian Updating. *Management Science*, 60(4):923–938.

Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2015). Money, Depletion, and Prosociality in the Dictator Game. *Journal of Neuroscience, Psychology, and Economics*, 8(1):1–14.

Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2016). The Impact of Self-Control Depletion on Social Preferences in the Ultimatum Game. *Journal of Economic Psychology*, 53:1–16.

Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2018). Social Preferences and Self-Control. *Journal of Behavioral and Experimental Economics*, 74:161–166.

Alger, I. and Weibull, J. W. (2013). Homo Moralis–Preference Evolution under Incomplete Information and Assortative Matching. *Econometrica*, 81(6):2269–2302.

Alós-Ferrer, C., Granić, D.-G., Kern, J., and Wagner, A. K. (2016). Preference Reversals: Time and Again. *Journal of Risk and Uncertainty*, 52(1):65–97.

Alós-Ferrer, C. and Ritschel, A. (2018). The Reinforcement Heuristic in Normal Form Games. *Journal of Economic Behavior and Organization*, forthcoming.

Alós-Ferrer, C. and Strack, F. (2014). From Dual Processes to Multiple Selves: Implications for Economic Behavior. *Journal of Economic Psychology*, 41:1–11.

Andersen, S., Ertaç, S., Gneezy, U., Hoffman, M., and List, J. A. (2011). Stakes Matter in Ultimatum Games. *American Economic Review*, 101(3):3427–3439.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, 97(6):1447–1458.

Andreoni, J. (1995). Warm-Glow Versus Cold-Prickle: The Effects of Positive and Negative Framing on Cooperation in Experiments. *Quarterly Journal of Economics*, 110(1):1–21.

Andreoni, J., Rao, J. M., and Trachtman, H. (2017). Avoiding the Ask: A Field Experiment on Altruism, Empathy, and Charitable Giving. *Journal of Political Economy*, 125(3):625–653.

Armantier, O. (2006). Do Wealth Differences Affect Fairness Considerations? *International Economic Review*, 47(2):391–429.

Azmat, G., Calsamiglia, C., and Iriberri, N. (2016). Gender Differences in Response to Big Stakes. *Journal of the European Economic Association*, 14(6).

Bardsley, N. (2008). Dictator Game Giving: Altruism or Artifact? *Experimental Economics*, 11(2):122–133.

Bellemare, C. and Kröger, S. (2007). On Representative Social Capital. *European Economic Review*, 51(1):183–202.

Berg, J., Dickhaut, J., and McCabe, K. (1995). Trust, Reciprocity, and Social History. *Games and Economic Behavior*, 10:122–142.

Blanco, M., Engelmann, D., and Normann, H. T. (2011). A Within-Subject Analysis of Other-regarding Preferences. *Games and Economic Behavior*, 72(2):321–338.

Bolton, G. E. and Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, 90(1):166–193.

Bosman, R., Hennig-Schmidt, H., and van Winden, F. (2006). Exploring Group Decision Making in a Power-to-Take Experiment. *Experimental Economics*, 9(1):35–51.

Bosman, R., Sutter, M., and van Winden, F. (2005). The Impact of Real Effort and Emotions in the Power-To-Take Game. *Journal of Economic Psychology*, 26(3):407–429.

Bosman, R. and van Winden, F. (2002). Emotional Hazard in a Power-to-Take Experiment. *Economic Journal*, 476(112):147–169.

Brandts, J. and Charness, G. (2011). The Strategy versus the Direct-response Method: A First Survey of Experimental Comparisons. *Experimental Economics*, 14(3):375–398.

Burks, S. V., Carpenter, J. P., and Verhoogen, E. (2003). Playing Both Roles in the Trust Game. *Journal of Economic Behavior and Organization*, 51(2):195–216.

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2016). Fairness is Intuitive. *Experimental Economics*, 19(4):727–740.

Cappelletti, D., Güth, W., and Ploner, M. (2011). Being of Two Minds: Ultimatum Offers under Cognitive Constraints. *Journal of Economic Psychology*, 32(6):940–950.

Carlsmith, J. M. and Gross, A. E. (1969). Some Effects of Guilt on Compliance. *Journal of Personality and Social Psychology*, 11(3):232.

Casari, M. and Cason, T. N. (2009). The Strategy Method Lowers Measured Trustworthy Behavior. *Economics Letters*, 103(3):157–159.

Charness, G. and Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, 117(3):817–869.

Croson, R. and Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2):448–474.

Dariel, A. and Nikiforakis, N. (2014). Cooperators and Reciprocators: A Within-subject Analysis of Pro-social behavior. *Economics Letters*, 122(2):163–166.

Dashiell, J. F. (1937). Affective Value-Distances as a Determinant of Aesthetic Judgment-Times. *American Journal of Psychology*, 50:57–67.

Declerck, C. H., Kiyonari, T., and Boone, C. (2009). Why Do Responders Reject Unequal Offers in the Ultimatum Game? An Experimental Study on the Role of Perceiving Interdependence. *Journal of Economic Psychology*, 30(3):335–343.

DellaVigna, S., List, J., and Malmendier, U. (2012). Testing for Altruism and Social Pressure in Charitable Giving. *Quarterly Journal of Economics*, 127(1):1–56.

Dufwenberg, M. and Kirchsteiger, G. (2004). A Theory of Sequential Reciprocity. *Games and Economic Behavior*, 47(2):268–298.

Eckel, C. C. and Petrie, R. (2011). Face Value. *American Economic Review*, 101(4):1497–1513.

Engel, C. (2011). Dictator Games: A Meta Study. *Experimental Economics*, 14(4):583–610.

Engelmann, D. and Strobel, M. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *American Economic Review*, 94(4):857–869.

Falk, A. and Fischbacher, U. (2006). A Theory of Reciprocity. *Games and Economic Behavior*, 54(2):293–315.

Fehr, E. and Schmidt, K. M. (1999). A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics*, 114(3):817–868.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-Made Economic Experiments. *Experimental Economics*, 10(2):171–178.

Fisman, R., Kariv, S., and Markovits, D. (2007). Individual Preferences for Giving. *American Economic Review*, 97(5):1858–1876.

Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in Simple Bargaining Experiments. *Games and Economic Behavior*, 6(3):347–369.

Galizzi, M. M. and Navarro-Martínez, D. (2018). On the External Validity of Social-Preference Games: A Systematic Lab-Field Study. *Management Science*, forthcoming.

Gilligan, C. (1982). *In a Different Voice*. Harvard University Press.

Giner-Sorolla, R. (2001). Guilty Pleasures and Grim Necessities: Affective Attitudes in Dilemmas of Self-Control. *Journal of Personality and Social Psychology*, 80(2):206.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A., and Soutter, C. L. (2000). Measuring Trust. *Quarterly Journal of Economics*, 115(3):811–846.

Gneezy, U., Imas, A., and Madarász, K. (2014). Conscience Accounting: Emotion Dynamics and Social Behavior. *Management Science*, 60(11):2645–2658.

Greiner, B. (2015). Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE. *Journal of the Economic Science Association*, 1:114–125.

Güth, W., Schmittberger, R., and Schwarze, B. (1982). An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization*, 3(4):367–388.

Johnson, N. D. and Mislin, A. A. (2011). Trust Games: A Meta-Analysis. *Journal of Economic Psychology*, 32(5):865–889.

Karakostas, A. and Zizzo, D. J. (2016). Compliance and the Power of Authority. *Journal of Economic Behavior and Organization*, 124:67–80.

Khadjavi, M. and Lange, A. (2015). Doing Good or Doing Harm: Experimental Evidence On Giving And Taking In Public Good Games. *Experimental Economics*, 18:432–441.

Krajbich, I., Bartling, B., Hare, T., and Fehr, E. (2015). Rethinking Fast and Slow Based on a Critique of Reaction-Time Reverse Inference. *Nature Communications*, 6:7455.

Krugman, P. (2008). Crisis of Confidence. *The New York Times*, April 14th.

Levitt, S. D. and List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, pages 153–174.

Lindsey, L. L. M. (2005). Anticipated Guilt as Behavioral Motivation: An Examination of Appeals to Help Unknown Others Through Bone Marrow Donation. *Human Communication Research*, 31(4):453–481.

List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3):482–493.

Malinowski, C. I. and Smith, C. P. (1985). Moral Reasoning and Moral Conduct: An Investigation Prompted by Kohlberg's Theory. *Journal of Personality and Social Psychology*, 49(4):1016–1027.

Mazar, N., Amir, O., and Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*, 45(6):633–644.

Mazar, N. and Zhong, C.-B. (2010). Do Green Products Make Us Better People? *Psychological Science*, 21:494–498.

McLeish, K. N. and Oxoby, R. J. (2011). Social Interactions and the Salience of Social Identity. *Journal of Economic Psychology*, 32(1):172–178.

Monin, B. and Miller, D. T. (2001). Moral Credentials and the Expression of Prejudice. *Journal of Personality and Social Psychology*, 81(1):33.

Mosteller, F. and Nogee, P. (1951). An Experimental Measurement of Utility. *Journal of Political Economy*, 59:371–404.

Moyer, R. S. and Landauer, T. K. (1967). Time Required for Judgements of Numerical Inequality. *Nature*, 215(5109):1519–1520.

Myerson, R. B. (2012). A Model of Moral-Hazard Credit Cycles. *Journal of Political Economy*, 120(5):847–878.

Myrseth, K. O. R. and Wollbrant, C. E. (2016). Commentary: Fairness is Intuitive. *Frontiers in Psychology*, 7(654).

Oosterbeek, H., Sloof, R., and van de Kuilen, G. (2004). Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis. *Experimental Economics*, 7(2):171–188.

Owens, L. A. (2012). Confidence in Banks, Financial Institutions and Wall Street, 1971–2011. *Public Opinion Quarterly*, 76:142–162.

Polivy, J. and Herman, C. P. (1985). Dieting and Binging: A Causal Analysis. *American Psychologist*, 40(2):193–201.

Reuben, E. and van Winden, F. (2010). Fairness Perceptions and Prosocial Emotions in the Power to Take. *Journal of Economic Psychology*, 31(6):908–922.

Roberts, T.-A. and Nolen-Hoeksema, S. (1989). Sex Differences in Reactions to Evaluative Feedback. *Sex Roles*, 21(11-12):725–747.

Roth, A. E., Prasnikar, V., Okuno-Fujiwara, M., and Zamir, S. (1991). Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study. *American Economic Review*, 81(5):1068–1095.

Schmidt, H. (2003). Das Gesetz des Dschungels [The Law of the Jungle]. *Die Zeit*, December 4th(50/2003).

Selten, R. (1967). Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In Sauermann, H., editor, *Beiträge zur experimentellen Wirtschaftsforschung*, pages 136–168. J. C. B. Mohr (Paul Siebeck), Tübingen.

Slonim, R. and Roth, A. E. (1998). Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic. *Econometrica*, 63(3):569–596.

Smith, A. (1759). *The Theory of Moral Sentiments*. Reprinted 1981 by Liberty Fund (D. D. Raphael and A. L. Macfie, editors), Indianapolis.

Stoop, J., Noussair, C. N., and van Soest, D. (2012). From the Lab to the Field: Cooperation among Fishermen. *Journal of Political Economy*, 120(6):1027–1056.

Walkowitz, G., Hennig-Schmidt, H., and Oberhammer, C. (2009). Experimenting over a Long Distance–A Method to Facilitate Inter-Cultural Experiments and its Application to a Trust Game. Working Paper, Department of Economics, University of Bonn.

Zemack-Rugar, Y., Bettman, J. R., and Fitzsimons, G. J. (2007). The Effects of Nonconsciously Priming Emotion Concepts on Behavior. *Journal of Personality and Social Psychology*, 93(6):927.

Zhong, C.-B., Ku, G., Lount, R. B., and Murnighan, J. K. (2010). Compensatory Ethics. *Journal of Business Ethics*, 92(3):323–339.

Zizzo, D. J. (2004). Inequality and Procedural Fairness in a Money Burning and Stealing Experiment. *Research on Economic Inequality*, 11:215–247.

Zizzo, D. J. and Oswald, A. J. (2001). Are People Willing to Pay to Reduce Others' Incomes? *Annales d'Economie et de Statistique*, (63–64):39–65.